

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO
Programa de Pós Graduação em Economia e Administração FEA – PUC-SP



Núcleo de Estudos do Futuro



PUC . SP . Brasil



GPS

**TEMA 11. BIODIVERSIDADE - CAPITAL
NATURAL E SOCIAL**

1. INTRODUÇÃO

O presente trabalho tem por objetivo analisar um conjunto de dados disponibilizados pelo Banco Mundial relativos ao tema **Biodiversidade - Capital Natural e Social**, que possui um conjunto de 33 (trinta e três) variáveis previamente selecionadas e que refletem o grau de engajamento sobre esta temática, considerando os 132 países objetos deste estudo.

Iniciamos o trabalho apresentando os objetivos do programa Países Sustentáveis – Gestão Pública Sustentável (GPS), uma publicação do Núcleo de Estudos do Futuro da PUC-SP. Em seguida algumas definições e contextualização sobre o que se entende por Biodiversidade, Capital Natural e Social.

Depois partimos para o entendimento dos dados apresentando resultados de análise exploratória dos dados sobre o tema, empregando os *softwares* estatísticos MINITAB e SPSS, Listar principais análises estatísticas utilizadas nas 12 sessões...

Por fim, apresentamos os resultados e considerações....

O objetivo é explorar os dados sobre **Biodiversidade: Capital Natural e Social** de tal modo que possamos com os recursos citados anteriormente apresentar conclusões de correlações entre as variáveis que compõem o tema.

1.1. ENTENDENDO OS DADOS

Antes da apresentação dos indivíduos desta análise que envolve 132 países e os indicadores de desenvolvimento humano selecionados a partir de referência internacional, citamos a origem do trabalho, que é o **Projeto ORIBER – Observatório RIBER**

1.1.1 O tema de pesquisa: Biodiversidade - Capital Natural e Social

A análise dos conceitos que compõem o tema foi extraída do material de aula e do Guia GPS. Nesse estudo são apresentadas as informações teóricas de outros indicadores, cuja análise e interpretação estão sob responsabilidade de outros colegas desse curso.

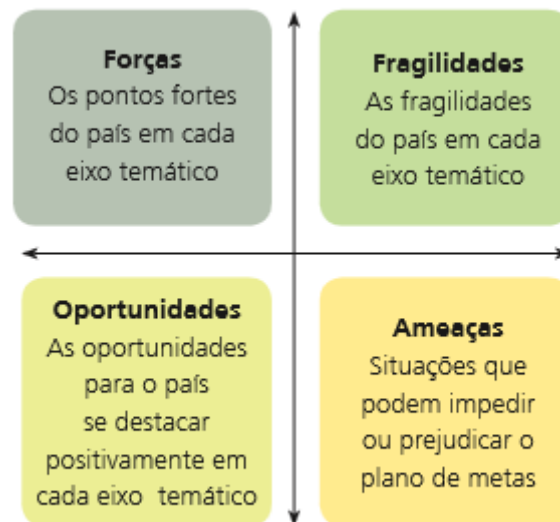
Vale destacar o desafio citado no GPS e sua proposta de solução:

Desafio: “A grande maioria dos países Ibero-americanos se depara com o seguinte desafio: - como equilibrar a necessária dinâmica econômica com a sustentabilidade ambiental e o equilíbrio social, no contexto de uma gestão aberta, democrática e participativa ?”

Solução: “O GPS - Guia para Gestão Pública Sustentável propõe então uma forma de contribuir na superação desses desafios, que consiste na promoção, a partir das secretarias de planejamento de cada país, de sinergias entre os setores científico-tecnológico, sociocultural e institucional, que harmonizem os processos e impactos do desenvolvimento ao nível local, tornando-o sustentável, procurando sempre estimular a participação dos cidadãos como forma de contribuir para a melhoria da qualidade de vida, e aproveitando de modo efetivo a troca de informações e experiências com outros países da região ibero-americana e suas redes de contatos.”

Para chegar a uma conclusão do estágio em que se encontra o país e como é possível adotar uma solução, o GPS propõe um levantamento dos dados a partir da análise SWOT, conforme se apresenta na figura 1.

Figura 1: Análise SWOT de cada país

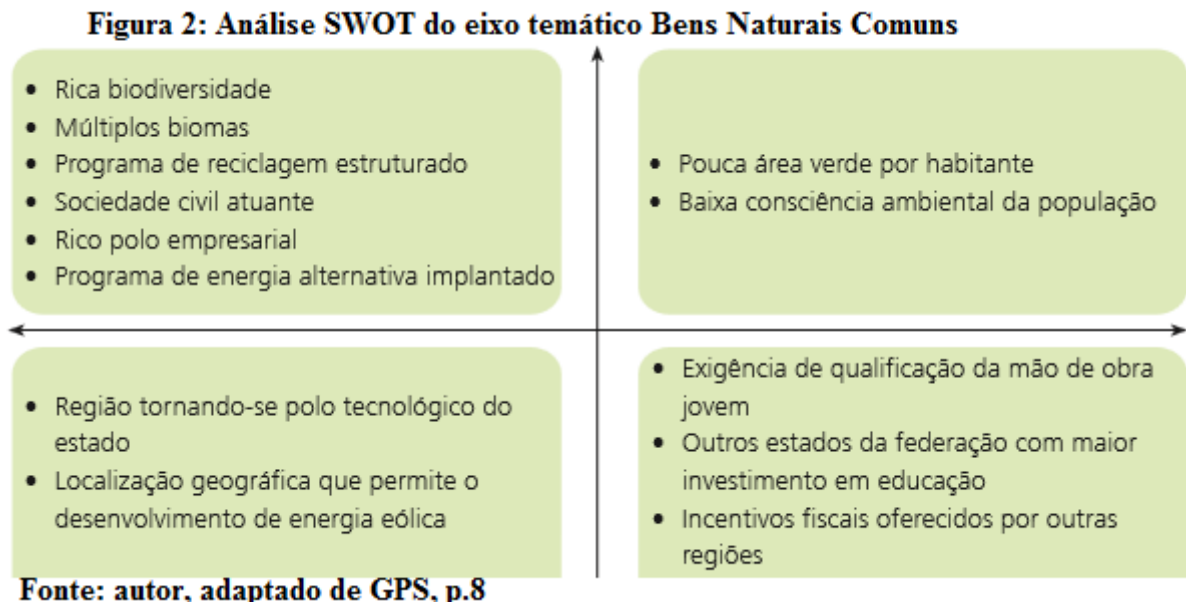


fonte: autor, adaptado de GPS, p.8

Na figura 1 é possível obter informações sobre o país analisado considerando suas Forças – quais são os seus pontos fortes em cada um dos eixos temáticos e suas Fragilidades, Ameaças e Oportunidades que cada eixo temático apresenta para o país em questão.

Como exemplo inicial deste trabalho referente ao tema: **Biodiversidade - Capital Natural e Social**, podemos destacar o “Alimento”. Supondo que o país analisado seja um grande produtor de alimento (ponto forte), tenha uma estiagem em seu território (fragilidade) e perceba uma ameaça (outro país produtor) mas identifica uma oportunidade, a de planejar a produção de certo alimento usando sistema de irrigação por gotejamento a partir do armazenamento e do uso racional desse recurso.

Para ajudar o leitor a entender a contribuição da matriz SWOT, o GPS apresenta um exemplo com o tema “Bens naturais comuns”, conforme mostra a figura 2.



Se tomarmos como ponto de partida o item do quadrante do ponto forte “Rica Biodiversidade” e avançarmos no sentido horário da matriz encontraremos a relação com “Pouca área verde por habitante”, que indica a fragilidade. Aparece como ameaça a “Exigência de qualificação da mão de obra jovem” e em seguida como oportunidade o fato de a “Região se tornar um polo tecnológico do estado”. Note que as variáveis vão sendo encontradas e analisadas como proposta de solução para o problema do tema daquele país.

Projeto Milênio – Observatório RIBER - ORIBER

A Rede Ibero-americana do Projeto Milênio – RIBER, tem como objetivo promover o estudo da prospectiva na Iberoamérica através de uma comunidade com competências, conceitos, métodos e práticas prospectivas, que permita analisar e intercambiar informações sobre os diferentes países da região e contribuir na aprendizagem coletiva dos atores regionais para o desenvolvimento de competências em prospectiva, fornecendo referências das melhores práticas existentes na região e no mundo como um todo.

Fonte: Guia GPS – Gestão Pública Sustentável

1.1.2 Os indivíduos

Os indivíduos desta análise são os 132 países analisados por uma série de indicadores de referência internacional. Os dados analisados de cada país são as variáveis que descrevemos a seguir.

1.1.2.1 As variáveis do tema: Biodiversidade - Capital Natural e Social

Este tema da pesquisa apresenta 33 (trinta e três) variáveis, sendo três categóricas – País, Country code e REGIÃO e e trinta quantitativas, descritas abaixo no Quadro 1. Estas ainda podem ser divididas em: índices sintéticos e variáveis-componentes ou indicadores, a saber:

Índices sintéticos: São 16 (dezesesseis): Índice de Progresso Social - Social Progress Index, Índice de Desenvolvimento Humano - IDH – 2013, Índice de Governança, Índice de Proteção Ambiental - EPI Score, EV - Water Resources – tratamento de águas residuais, EV Agriculture, EV Forests, EV Fisheries - Ecosystem Vitality (índice de vitalidade do ecossistema), Basic Human Needs, Foundations of Wellbeing, Opportunity, Health and Wellness, Personal Freedom and Choice, Índice de Felicidade - Happy Planet Index, Índice de GINI e Índice de Democracia - Democracy Index – 2014.

Indicadores ou variáveis componentes: Inclui 14 variáveis como: Satisfied demand for contraception (% of women), Greenhouse gas emissions (CO2 equivalents per GDP) - (4=low; 0=high), Water withdrawals as a percent of resources, Biodiversity and habitat (0=no protection; 100=high protection), Modern slavery, human trafficking and child marriage (1=low; 100=high), Women treated with respect (0=low; 100=high), Tolerance for immigrants (0=low; 100=high), Tolerance for homosexuals (0=low; 100=high), Discrimination and violence against minorities (0=low; 10=high), Religious tolerance (1=low; 4=high), Community safety net (0=low; 100=high), Years of tertiary schooling, Women's average years in school e Inequality in the attainment of education (0=low; 1=high)

Quadro 1. As 30 variáveis do eixo temático: Biodiversidade - Capital Natural e Social

Variável	Significado	Tipo	Unidade de Medida
Social Progress Index	O Índice de Progresso Social (Social Progress Index), é composto por três dimensões do progresso social: Necessidades Humanas Básicas, Fundamentos de Bem-Estar e Oportunidades. fonte: www.socialprogressimperative.org	Variável Quantitativa	Indicador nominal
Basic Human Needs	Necessidades básicas do ser humano (alimentação básica e cuidados médicos). É medido por uma escala de 1 a 100, sendo: 1 para baixa atenção e 100 para alta atenção às necessidades básicas do ser humano.	Variável Quantitativa	Indicador nominal
Foundations of Wellbeing	Fundamentos do Bem Estar e melhoria de vida através da educação	Variável Quantitativa	Indicador nominal
Opportunity	Indicador composto por: Direitos individuais, Liberdades individuais, Tolerância e inclusão e Acesso à educação superior	Variável Quantitativa	Indicador nominal
Health and Wellness	Qualidade de vida, através de melhores condições de saúde e de bem estar	Variável Quantitativa	Indicador nominal

Personal Freedom and Choice	Liberdade na tomada de decisão e escolhas de vida do indivíduo e da população	Variável Quantitativa	Percentual
Greenhouse gas emissions (CO2 equivalents per GDP)	Emissão de poluentes na atmosfera. Escala de 0 a 4, sendo: 4: < 100, 3: 100 – 200, 2: 200 – 1000, 1: 1000 – 2000 e 0: > 2000	Variável Quantitativa	Indicador nominal que varia de 0 a 4
Water withdrawals as a percent of resources	Uso de água como porcentagem dos recursos. Obedece uma escala de 0 a 5, sendo: 4–5: Extremely high stress (>80%), 3–4: High stress (40–80%), 2–3: Medium-high stress (20–40%), 1–2: Low-medium stress (10–20%) e 0–1: Low stress (<10%)	Variável Quantitativa	Indicador nominal que varia de 0 a 5
Biodiversity and habitat (0=no protection; 100=high protection)	Variável composta por componentes que retratam o nível de proteção das áreas terrestres e marinhas, bem como as espécies ameaçadas ou em vias de extinção e seus biomas.	Variável Quantitativa	Indicador nominal que varia de 1 a 100
Modern slavery, human trafficking and child marriage (1=low; 100=high)	Esta variável é a combinação de 3 variáveis que medem: a escravidão nos tempos atuais (privação da liberdade e trabalho forçado), o tráfico de pessoas e exploração, e o casamento forçado de indivíduos menores de 18 anos.	Variável Quantitativa	Indicador nominal que varia de 1 a 100
Satisfied demand for contraception (% of women)	Variável que mede o grau de satisfação de mulheres com idade entre 15 e 49 anos, satisfeitas com os métodos contraceptivos disponíveis.	Variável Quantitativa	Percentual
Women treated with respect (0=low; 100=high)	Variável que contabiliza respostas SIM (positivas) de mulheres respondentes sobre sua percepção sobre serem tratadas com respeito e dignidade em seus países.	Variável Quantitativa	Percentual
Tolerance for immigrants (0=low; 100=high)	Variável que contabiliza respostas SIM (positivas) da população sobre a percepção de onde vivem ser um bom lugar para imigrantes de outros países.	Variável Quantitativa	Percentual
Tolerance for homosexuals (0=low; 100=high)	Variável que contabiliza respostas SIM (positivas) da população sobre a percepção de onde vivem ser um bom lugar para a comunidade LGBT (gays, lésbicas,...).	Variável Quantitativa	Percentual
Discrimination and violence against minorities (0=low; 10=high)	Variável que mede o grau de intolerância, discriminação e violência contra as minorias e desfavorecidos (éticas, religiosas,...). É medida por uma escala de 0 a 10, sendo: 0 (low pressures) to 10 (very high pressures).	Variável Quantitativa	Indicador nominal que varia de 0 a 10
Religious tolerance (1=low; 4=very high)	Esta variável é a medida de 13 tipos de hostilidades causadas por indivíduos, organizações ou sociedade (assédio, abuso, intimidação,..)	Variável Quantitativa	Indicador nominal que varia de 1 a 4

Community safety net (0=low; 100=high)	Variável que contabiliza respostas SIM (positivas) da população sobre uma situação problemática, onde possam contar com seus familiares ou amigos onde estiverem.	Variável Quantitativa	Percentual
Years of tertiary schooling	Média de anos para completar o terceiro grau (ensino superior) para pessoas com idade superior a 25 anos de idade. Quanto maior o indicador, melhor o resultado do país.	Variável Quantitativa	Indicador nominal (em anos)
Women's average years in school	Média de anos para mulheres que frequentaram a escola, com idade entre 25 e 34 anos. Incluindo primeiro, segundo e terceiro grau	Variável Quantitativa	Indicador nominal (em anos)
Inequality in the attainment of education (0=low; 1=high)	Variável mede a perda no aprendizado devido a falta de qualidade no ensino. Escala de 0, para baixa perda e 1 para alta perda.	Variável Quantitativa	Indicador nominal que varia de 0 a 1
IDH - 2013	O Índice de Desenvolvimento Humano (IDH) mede o progresso de uma nação a partir de três dimensões: renda, saúde e educação.	Variável Quantitativa	Indicador nominal
1. Governança	Indicador que mede a capacidade de um país de eleger, monitorar e substituir seus governantes eleitos. Mede também nível de estabilidade política, medidas contra corrupção, entre outros.	Variável Quantitativa	Indicador nominal
EPI Score	Índice de Performance Ambiental, composto por: Saúde Ambiental (saúde humana e dos ecossistemas), Impactos na Saúde (medido pela mortalidade infantil – idade entre 1 e 5 anos), Qualidade do Ar (poluição do ar, partículas sólidas e seu impacto na saúde) e Nível de Água tratada e Sanitização adequada.	Variável Quantitativa	Indicador nominal
EV - Water Resources	Wastewater treatment level weighted by connection to wastewater treatment rate.	Variável Quantitativa	Indicador nominal
EV - Agriculture	Variável que mede os subsídios destinados à Agricultura e a Regulamentação no uso de pesticidas conforme tratado assinado na Convenção de Estocolmo.	Variável Quantitativa	Indicador nominal
EV - Forests	Variável que demonstra a variação das áreas cobertas por floresta (desmatamento, queima e exploração).	Variável Quantitativa	Indicador nominal
EV - Fisheries	Pesca costeira intensiva com a utilização de equipamentos de arrasto e dragagem (medido em ton métrica), e nível de estocagem do excedente produzido. Quanto maior o índice, mais intensiva é a prática. (Coastal Shelf Fishing Pressure and Fish Stocks).	Variável Quantitativa	Indicador nominal
Happy Planet Index	HPI, Índice de Felicidade do Planeta, medida do Bem Estar Sustentável (Longevidade, Felicidade e Vida sustentável percebida pela população do país).	Variável Quantitativa	Indicador nominal
GINI Index	Medida de desigualdade social utilizada para medir a desigualdade na distribuição de renda, ou desigualdade da riqueza. Valores mais baixos	Variável Quantitativa	Indicador nominal

	representam menor desigualdade.		
Democracy Index - 2014	A variável Índice de Democracia é formada por 5 categorias: Processo eleitoral e pluralismo, Liberdade civil, Funcionamento do governo, Participação política e Cultura política.	Variável Quantitativa	Indicador nominal

Fonte: autor a partir dos dados da planilha estatística e do GPS p.16

1.1.2.2 A Tabela de Dados

Em função da extensão da tabela, vamos representar aqui **apenas alguns países e algumas** variáveis selecionadas para ilustrar as definições do Quadro 1.

Tabela 1: Países e variáveis selecionadas

PAÍS (132)	Country Code	1	48	49	50	58	59	61	62	63
		Social Progress Index	Greenhouse gas emissions (CO2 equivalents per GDP)	Water withdrawals as a percent of resources	Biodiversity and habitat (0=no protection; 100=high protection)	Modern slavery, human trafficking and child marriage (1=low; 100=high)	Satisfied demand for contraception (% of women)	Women treated with respect (0=low; 100=high)	Tolerance for immigrants (0=low; 100=high)	Tolerance for homosexuals (0=low; 100=high)
Albania	ALB	69.13	267.57	1.90	63.19	11.07	82.40	44.00	45.00	12.00
Algeria	DZA	59.13	629.65	3.44	34.86	7.29	82.50	49.00	49.00	21.00
Angola	AGO	39.93	2227.70	1.54	41.70	4.84	31.20	62.00	41.00	15.00
Argentina	ARG	70.59	618.53	2.51	44.88	4.35	86.90	26.00	72.00	64.00
Armenia	ARM	65.03	886.10	3.07	80.16	10.57	74.50	68.00	52.00	3.00
Australia	AUS	86.10	769.09	3.51	83.08	2.15	87.00	83.00	87.00	72.00
Austria	AUT	85.11	284.46	0.32	86.59	1.14	80.40	78.00	63.00	48.00
Azerbaijan	AZE	62.44	795.71	3.39	58.66	11.07	78.60	77.00	50.00	6.00
Bangladesh	BGD	52.04	588.57	0.65	39.68	9.54	82.40	81.00	57.00	8.00
Belarus	BLR	65.20	816.25	0.59	41.25	4.82	86.50	67.00	46.00	10.00
Belgium	BEL	82.63	439.90	3.16	57.56	1.14	89.90	77.00	64.00	67.00
Benin	BEN	49.11	1370.12	0.00	65.88	23.57	40.60	65.00	77.00	4.00
Bolivia	BOL	62.90	1387.13	0.68	81.04	9.30	75.10	34.00	69.00	37.00
Bosnia and Herzegovina	BIH	64.99	1021.09	0.02	2.50	10.45	70.60	42.00	51.00	10.00
Botswana	BWA	65.60	458.01	1.36	100.00	19.92	74.00	71.00	60.00	24.00
Brazil	BRA	69.97	590.89	0.91	66.74	7.16	91.50	28.00	69.00	62.00
Bulgaria	BGR	70.24	713.96	1.27	68.85	10.80	85.50	61.00	46.00	29.00
Burkina Faso	BFA	47.33	1088.93	0.00	83.75	20.65	35.80	67.00	86.00	5.00
Burundi	BDI	37.33	8442.23	0.00	30.29	19.64	43.00	52.00	57.00	3.00
Cambodia	KHM	51.89	934.71	0.44	78.93	19.99	67.60	92.00	25.00	22.00
Cameroon	CMR	45.51	2241.41	0.11	45.67	20.87	53.50	62.00	74.00	8.00
Canada	CAN	86.95	604.51	1.21	58.40	1.73	90.20	85.00	93.00	81.00

Central African Republic	CAF	34.17	28763.64	0.00	97.00	20.36	53.40	61.00	74.00	11.00
Chad	TCD	32.60	2523.22	0.52	55.34	20.19	20.30	36.00	61.00	4.00
Chile	CHL	76.30	418.10	3.21	60.16	7.44	82.20	37.00	68.00	58.00
China	CHN	58.67	1138.49	2.94	66.63	8.59	96.20	89.00	45.00	14.00
Colombia	COL	67.24	443.28	0.33	80.40	9.50	90.40	17.00	68.00	47.00
Congo, Republic of	COG	47.99	976.94	0.43	66.29	21.01	69.10	54.00	72.00	9.00
Costa Rica	CRI	77.75	308.71	1.94	89.55	1.66	94.50	37.00	77.00	59.00
Croatia	HRV	73.31	380.19	0.02	70.87	10.32	85.10	47.00	35.00	22.00
Cuba	CUB	61.07	471.29	3.19	56.06	1.28	88.30		50.00	54.00
Czech Republic	CZE	80.41	469.03	1.13	98.35	10.82	87.60	70.00	49.00	43.00
Denmark	DNK	86.55	322.54	0.95	67.74	1.14	88.90	91.00	77.00	67.00
Djibouti	DJI	45.95	634.45	3.39	15.45	10.00	44.30	84.00	56.00	5.00
Dominican Republic	DOM	63.03	365.06	3.49	92.94	9.01	84.90	18.00	63.00	34.00
Ecuador	ECU	68.15	461.80	1.86	88.09	9.78	88.90	43.00	60.00	36.00
Egypt	EGY	59.97	639.20	1.33	65.26	5.09	84.30	69.00	25.00	
El Salvador	SLV	64.70	331.03	1.49	24.53	7.12	89.30	27.00	61.00	50.00
Estonia	EST	81.28	992.42	2.75	100.00	3.47	84.30	61.00	35.00	27.00
Finland	FIN	86.91	413.48	0.98	61.86	1.14	90.70	84.00	72.00	66.00
France	FRA	81.11	283.98	1.75	54.45	2.13	93.00	69.00	66.00	62.00
Georgia	GEO	63.94	691.05	1.51	29.59	11.56	73.90	81.00	47.00	6.00
Germany	DEU	84.61	337.60	1.90	100.00	2.63	89.00	78.00	75.00	63.00
Ghana	GHA	55.96	724.48	0.11	34.26	21.04	40.30	69.00	59.00	4.00
Greece	GRC	73.43	434.13	3.27	66.49	1.14	86.50	64.00	40.00	21.00
Guatemala	GTM	61.37	566.93	1.01	57.55	5.75	69.30	26.00	45.00	22.00
Guinea	GIN	37.41	2030.19	0.06	41.00	21.26	30.50	57.00	74.00	6.00
Guyana	GUY	60.06	1833.51	1.78	18.66	9.29	58.10		56.00	19.00
Honduras	HND	61.28	705.68	0.07	76.81	5.42	81.70	35.00	56.00	32.00
Hungary	HUN	73.87	409.82	0.49	29.57	10.82	90.70	58.00	53.00	24.00
Iceland	ISL	88.07	407.88	0.14	81.90	1.00		80.00	78.00	79.00
India	IND	50.24	618.09	3.58	39.18	30.84	80.00	61.00	30.00	11.00
Indonesia	IDN	58.98	883.43	3.26	78.08	5.08	82.10	66.00	22.00	2.00
Iran	IRN	56.65	934.71	4.78	37.06	5.70	89.60	65.00	50.00	
Iraq	IRQ	44.84	2189.29	3.48	0.00	5.61	71.40	30.00	37.00	10.00
Ireland	IRL	84.05	374.49	2.92	18.74	1.00	85.30	88.00	84.00	75.00
Israel	ISR	71.40	447.89	4.83	39.34	5.23	88.60	61.00	32.00	31.00
Italy	ITA	76.93	314.44	3.35	79.77	3.12	84.40	64.00	70.00	56.00
Jamaica	JAM	70.39	680.86	5.00	88.04	4.03	86.50	29.00	74.00	15.00
Japan	JPN	84.21	328.44	3.05	73.53	3.31	75.90	62.00	58.00	34.00
Jordan	JOR	61.92	810.02	4.59	8.52	7.87	80.80	74.00	31.00	
Kazakhstan	KAZ	59.47	1376.80	4.02	14.11	8.51	79.80	73.00	52.00	7.00
Kenya	KEN	50.20	771.43	0.68	67.96	5.72	65.60	48.00	54.00	12.00
Korea, Republic of	KOR	77.18	512.75	3.54	50.40	2.32	93.40	55.00	56.00	14.00
Kuwait	KWT	70.66	1666.98	4.96	92.88	6.63	76.00	92.00	69.00	
Kyrgyzstan	KGZ	57.08	994.05	4.82	15.77	8.73	80.20	59.00	47.00	5.00
Laos	LAO	52.41	1490.45	0.01	93.85	20.38	66.70	88.00	63.00	41.00

Latvia	LVA	73.91	422.72	0.35	95.17	3.21	85.50	56.00	34.00	20.00
Lebanon	LBN	60.05	424.72	4.54	1.71	4.72	81.40	60.00	53.00	7.00
Lesotho	LSO	48.94	850.05	3.97	1.30	19.34	67.10	61.00	44.00	18.00
Liberia	LBR	44.02	833.81	0.27	4.13	20.32	27.00	72.00	71.00	17.00
Lithuania	LTU	73.76	447.46	1.19	91.83	3.63	79.50	48.00	36.00	16.00
Macedonia	MKD	68.33	593.75	2.65	27.54	8.94	82.60	64.00	51.00	12.00
Madagascar	MDG	44.28	2699.67	1.25	37.08	5.22	66.30	38.00	54.00	9.00
Malawi	MWI	48.79	1491.23	0.11	87.81	20.20	62.70	62.00	55.00	5.00
Malaysia	MYS	70.00	738.12	2.09	93.37	4.48	76.50	82.00	20.00	6.00
Mali	MLI	46.85	1684.87	0.20	14.30	20.07	24.20	84.00	80.00	2.00
Mauritania	MRT	43.11	1191.34	0.60	32.63	97.90	28.10	79.00	54.00	10.00
Mauritius	MUS	73.68	397.30		43.70	1.81	92.30	73.00	79.00	24.00
Mexico	MEX	66.41	487.36	3.52	62.32	5.57	86.90	51.00	42.00	39.00
Moldova	MDA	60.12	1083.18	1.46	7.51	25.68	85.30	34.00	41.00	9.00
Mongolia	MNG	58.97	2354.61	4.05	73.25	5.45	81.60	50.00	38.00	14.00
Montenegro	MNE	66.80	6220.00	0.31	51.60	10.94	72.60	45.00	54.00	15.00
Morocco	MAR	58.01	322.89	4.24	31.20	7.23	85.60	56.00	47.00	7.00
Mozambique	MOZ	45.23	1356.28	0.82	63.17	20.07	35.90	40.00	77.00	38.00
Namibia	NAM	61.19	1185.09	1.88	94.39	18.89	74.50		56.00	18.00
Nepal	NPL	51.58	1135.44	2.40	61.79	26.56	65.00	50.00	60.00	23.00
Netherlands	NLD	87.37	465.89	1.73	94.69	2.13	86.50	86.00	79.00	85.00
New Zealand	NZL	88.24	703.28	1.35	76.05	1.10	89.10	82.00	87.00	69.00
Nicaragua	NIC	62.33	920.17	1.01	91.67	5.47	87.50	38.00	64.00	57.00
Niger	NER	40.10	2242.32	0.11	77.21	20.50	39.00	78.00	72.00	3.00
Nigeria	NGA	42.65	915.58	0.29	57.54	13.99	40.40	51.00	62.00	7.00
Norway	NOR	87.12	243.66	0.40	71.68	1.14	93.70	86.00	74.00	66.00
Pakistan	PAK	42.40	753.36	4.31	41.43	32.11	55.50	69.00	38.00	30.00
Panama	PAN	72.58	525.57	0.42	86.96	1.67	73.40	47.00	53.00	44.00
Paraguay	PRY	62.65	1167.57	0.01	31.20	9.49	92.70	32.00	87.00	45.00
Peru	PER	66.29	320.51	3.20	70.36	10.04	91.30	24.00	61.00	34.00
Philippines	PHL	65.86	448.50	3.33	64.67	6.73	68.70	78.00	63.00	59.00
Poland	POL	77.44	568.80	1.31	91.76	10.32	87.50	63.00	39.00	19.00
Portugal	PRT	80.49	345.44	3.34	64.36	1.64	92.80	58.00	75.00	48.00
Romania	ROU	67.72	559.77	0.84	63.51	4.00	88.00	35.00	40.00	17.00
Russia	RUS	60.79	1151.93	1.23	53.39	12.43	92.70	46.00	42.00	12.00
Rwanda	RWA	49.46	419.68	0.00	56.32	19.62	70.60	97.00	56.00	2.00
Saudi Arabia	SAU	64.38	799.83	4.99	93.70	8.73	56.50	89.00	69.00	
Senegal	SEN	53.52	986.77	0.21	85.36	22.31	27.70	62.00	77.00	3.00
Serbia	SRB	70.61	975.08	0.58	99.78	10.45	81.70	40.00	68.00	21.00
Slovakia	SVK	78.93	399.86	0.20	84.33	10.32	88.60	65.00	41.00	27.00
Slovenia	SVN	81.65	385.03	0.03	100.00	9.83	90.60	65.00	56.00	38.00
South Africa	ZAF	62.96	1176.39	3.04	63.96	4.99	83.40	54.00	56.00	47.00
Spain	ESP	80.77	329.19	3.73	56.56	1.64	83.50	69.00	81.00	80.00
Sri Lanka	LKA	59.71	411.41	3.01	67.70	4.81	90.80	72.00	44.00	4.00
Sudan	SDN	38.45	2084.31	0.91	16.29	20.86	29.00	60.00	53.00	7.00
Swaziland	SWZ	48.87	489.39	3.11	17.69	3.95	79.00	56.00	55.00	18.00

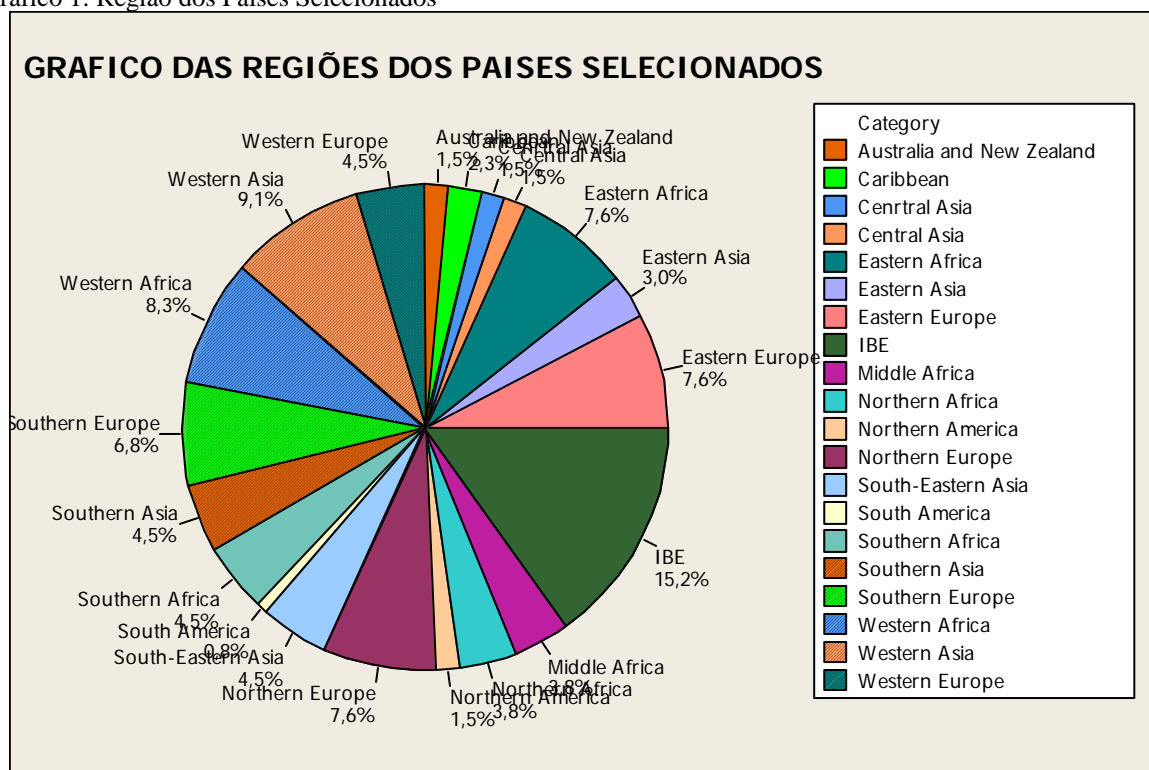
Sweden	SWE	87.08	231.19	1.30	62.38	1.14	88.30	73.00	80.00	69.00
Switzerland	CHE	88.19	170.50	1.06	100.00	1.14	92.10	90.00	70.00	65.00
Tajikistan	TJK	56.05	782.66	3.53	24.28	9.59	65.00	87.00	44.00	1.00
Tanzania	TZA	46.06	1292.02	1.50	79.12	20.47	57.30	55.00	43.00	16.00
Thailand	THA	65.14	720.14	1.70	70.19	20.78	93.80	80.00	23.00	27.00
Togo	TGO	42.80	1493.51	0.12	44.23	21.53	31.60	58.00	67.00	8.00
Trinidad and Tobago	TTO	69.88	1734.18	5.00	83.48	2.81	70.50	42.00	72.00	46.00
Tunisia	TUN	62.96	310.47	3.44	31.08	4.35	83.90	80.00	56.00	11.00
Turkey	TUR	64.62	422.04	3.02	32.62	7.40	89.30	38.00	47.00	14.00
Uganda	UGA	47.75	716.27	0.00	91.31	20.73	44.30	67.00	63.00	5.00
Ukraine	UKR	64.91	1411.36	2.10	41.46	8.42	86.30	44.00	38.00	14.00
United Arab Emirates	ARE	72.92	832.63	5.00	96.47	7.74	66.90	99.00	80.00	
United Kingdom	GBR	84.56	307.12	2.63	70.11	1.00	94.70	88.00	77.00	75.00
United States	USA	82.77	528.55	2.89	63.35	2.77	92.90	75.00	82.00	70.00
Uruguay	URY	77.51	817.95	0.86	24.62	9.51	91.50	47.00	82.00	76.00
Uzbekistan	UZB	57.34	2796.96	4.32	12.87	15.67	87.70	96.00	51.00	5.00
Venezuela	VEN	63.78	829.55	2.30	92.92	9.34	85.50	59.00	67.00	51.00
Yemen	YEM	40.23	507.26	4.67	23.59	7.28	53.40	68.00	25.00	
Zambia	ZMB	49.88	3734.91	0.08	100.00	19.83	63.10	61.00	68.00	5.00

Fonte: autor com base na tabela de dados da aula de Métodos Quantitativos

1.3. ANÁLISE DAS VARIÁVEIS

Apresentamos o gráfico circular ou gráfico de pizza (ou gráfico de torta) elaborado pelo software MINITAB com as regiões dos países selecionados.

Gráfico 1: Região dos Países Selecionados



1.3.1 – Variáveis Quantitativas

Vemos no Quadro 1, as variáveis e suas descrições. Essa informação aparece na Tabela 1 com uma seleção de países e indicadores com os números em percentual.

Com isso, a variável é denominada quantitativa pois mostra que ela pode ser medida em uma escala quantitativa, isto é, com números (SANTOS; PARRA FILHO, 2011).

Elas são Variáveis Contínuas porque assumem valores fracionados já que estão em percentual (IDH e GINI).

1.3.2 – Média, Mínimo e Máximo das variáveis com uso do MINITAB

Para facilitar o acompanhamento das análises, algumas definições se fazem necessárias para entender a apresentação dos dados do *software* MINITAB.

- Valor de N é o total de elementos da amostra, que neste caso é de: 132 países
- N* indica o total de dados ausentes na tabela em cada variável. Por exemplo, a variável EV – Fisheries, mostra um total de 37 países não divulgaram seus resultados.
- Mean (média) mostra o valor médio da variável para os 132 países.
- SE Mean mostra o erro padrão da amostra, isto é, ele mostra o quanto a média varia em relação as outras da amostra.
- StDev é o desvio-padrão mostra a distância estimada entre a média e os valores individuais da amostra. Quanto maior for o desvio padrão maior a dispersão ou distância dos dados em relação a média.
- Minimum mostra o valor mínimo dos dados dos 132 países
- Maximum mostra o valor máximo dos dados dos 132 países

Abaixo a tabela de estatística descritiva para os dados da base referente ao tema **Biodiversidade - Capital Natural e Social**.

Descriptive Statistics: Social Progr, Basic Human , Foundations , ...

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1
Social Progress Index	132	0	63.67	1.24	14.20	32.60	51.93
Basic Human Needs	132	0	70.00	1.63	18.77	25.94	55.14
Foundations of Wellbeing	132	0	67.529	0.994	11.424	41.340	59.465
Opportunity	132	0	53.48	1.35	15.56	24.31	42.38
Health and Wellness	132	0	72.458	0.606	6.958	49.930	68.225
Personal Freedom and Cho	132	0	67.52	1.12	12.87	17.62	59.93
Greenhouse gas emissions	132	0	1177	228	2623	170	427
Water withdrawals as a p	131	1	1.933	0.136	1.559	0.000	0.442
Biodiversity and habitat	132	0	59.51	2.43	27.90	0.00	37.60
Modern slavery, human tr	132	0	10.441	0.931	10.691	1.000	4.010
Satisfied demand for con	131	1	74.58	1.68	19.18	20.30	66.90
Women treated with respe	129	3	61.57	1.67	19.01	17.00	47.50
Tolerance for immigrants	132	0	57.82	1.42	16.32	20.00	46.00
Tolerance for homosexual	125	7	28.62	2.13	23.85	1.00	8.50
Discrimination and viole	132	0	5.990	0.176	2.019	1.000	4.450
Religious tolerance (1=1	132	0	2.7803	0.0856	0.9832	1.0000	2.0000
Community safety net (0=	132	0	79.42	1.10	12.63	30.00	73.00
Years of tertiary school	117	15	0.4475	0.0344	0.3717	0.0100	0.1400
Women`s average years in	131	1	9.471	0.327	3.743	1.300	6.600
Inequality in the attain	117	15	0.1960	0.0136	0.1474	0.0130	0.0555
IDH - 2013	125	7	0.6894	0.0144	0.1611	0.3370	0.5595
1. Governança	132	0	53.38	1.88	21.59	15.22	37.11
EPI Score	132	0	52.48	1.44	16.55	18.43	39.30
EV - Water Resources	132	0	27.77	2.85	32.71	0.00	0.54

EV - Agriculture	132	0	71.37	1.65	18.98	14.66	58.52
EV - Forests	110	22	39.72	3.02	31.65	0.00	14.76
EV - Fisheries	95	37	18.09	1.34	13.08	0.00	2.52
Happy Planet Index	129	3	42.673	0.796	9.040	22.591	36.163
GINNI Index	119	13	39.024	0.782	8.532	24.820	32.780
Democracy Index - 2014	132	0	5.873	0.180	2.068	1.490	4.100

Conforme observamos na tabela acima, identificamos dados faltantes (*missing values*) para algumas variáveis, que estão evidenciadas pelo valor diferente de 0 (zero) na coluna N*. Para fins deste estudo, as células que estiverem em branco serão preenchidas com o valor correspondente a média da variável em questão.

Após verificação na base de dados, as seguintes células foram substituídas para os respectivos países que apresentaram dados faltantes (*missing values*):

Water withdrawals as a percent of resources (1 caso) - Incluímos a média 1.933 no país: Mauritius

Satisfied demand for contraception (% of women) (1 caso) - Incluímos a média 74.58 no país: Iceland

Women treated with respect (3 casos) - Incluímos a média 61.57 nos países: Cuba, Guyana e Namibia

Tolerance for homosexuals (7 casos) - Incluímos a média 28.62 nos países: Egypt, Iran, Jordan, Kuwait, Saudi Arabia, United Arab Emirates e Yemen

Years of tertiary schooling (15 casos) - Incluímos a média 0.4475 nos países: Angola, Azerbaijan, Belarus, Bosnia and Herzegovina, Burkina Faso, Chad, Djibouti, Georgia, Guinea, Lebanon, Macedonia, Madagascar, Montenegro, Nigeria e Uzbekistan

Women's average years in school (1 caso) - Incluímos a média 9.471 no país: Iceland

Inequality in the attainment of education (15 casos) - Incluímos a média 0.1960 nos países: Algeria, Botswana, Burundi, Cuba, Iran, Japan, Kuwait, Malaysia, New Zealand, Paraguay, Russia, Saudi Arabia, Sudan, Tunisia e United Arab Emirates

IDH-2013 (7 casos) - Incluímos a média 0.6894 nos países: Algeria, Kazakhstan, Korea, Republic of Kyrgyzstan, Latvia, Mauritius e Netherlands

EV-Forests (22 casos) - Incluímos a média 39.72 nos países: Benin, Botswana, Burkina Faso, Chad, Djibouti, Egypt, Iceland, Iraq, Israel, Jordan, Kuwait, Lesotho, Mali, Mauritania, Namibia, Niger, Saudi Arabia, Senegal, Sudan, Tajikistan, United Arab Emirates e Yemen

EV - Fisheries (37 casos) - Incluímos a média 18.09 nos países: Armenia, Austria, Azerbaijan, Belarus, Bolivia, Bosnia and Herzegovina, Botswana, Burkina Faso, Burundi, Central African Republic, Chad, Czech Republic, Hungary, Iraq, Jordan, Kazakhstan, Kyrgyzstan, Laos, Lesotho, Macedonia, Malawi, Mali, Moldova, Mongolia, Nepal, Niger, Paraguay, Rwanda, Serbia, Slovakia, Slovenia, Swaziland, Switzerland, Tajikistan, Uganda, Uzbekistan e Zambia

Happy Planet Index (3 casos) - Incluímos a média 42.673 nos países: Lesotho, Montenegro e Swaziland

GINI Index (13 casos) - Incluímos a média 39.024 nos países: Bangladesh, Cuba, Guineia, Iceland, Kuwait, Lebanon, Mauritius, Montenegro, Nepal, Pakistan, Saudi Arabia, Sri Lanka e United Arab Emirates

Após a inclusão dos dados ausentes pela média de cada variável, foi executado novamente os indicadores de número de elementos (N), dados ausentes (N*), média, erro da média (SE mean), desvio padrão (Stdev), valor mínimo e máximo; apenas para confirmar se foram substituídos corretamente. Portanto, sem alteração nos demais valores.

Descriptive Statistics: Social Progr, Basic Human , Foundations , ...

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1
Social Progress Index	132	0	63.67	1.24	14.20	32.60	51.93
Basic Human Needs	132	0	70.00	1.63	18.77	25.94	55.14
Foundations of Wellbeing	132	0	67.529	0.994	11.424	41.340	59.465
Opportunity	132	0	53.48	1.35	15.56	24.31	42.38
Health and Wellness	132	0	72.458	0.606	6.958	49.930	68.225
Personal Freedom and Cho	132	0	67.52	1.12	12.87	17.62	59.93
Greenhouse gas emissions	132	0	1177	228	2623	170	427
Water withdrawals as a p	132	0	1.933	0.135	1.553	0.000	0.455
Biodiversity and habitat	132	0	59.51	2.43	27.90	0.00	37.60
Modern slavery, human tr	132	0	10.441	0.931	10.691	1.000	4.010
Satisfied demand for con	132	0	74.58	1.66	19.11	20.30	66.95
Women treated with respe	132	0	61.57	1.64	18.79	17.00	48.00
Tolerance for immigrants	132	0	57.82	1.42	16.32	20.00	46.00
Tolerance for homosexual	132	0	28.62	2.02	23.20	1.00	9.00
Discrimination and viole	132	0	5.990	0.176	2.019	1.000	4.450
Religious tolerance (1=1	132	0	2.7803	0.0856	0.9832	1.0000	2.0000
Community safety net (0=	132	0	79.42	1.10	12.63	30.00	73.00
Years of tertiary school	132	0	0.4475	0.0304	0.3498	0.0100	0.1525
Women`s average years in	132	0	9.471	0.325	3.728	1.300	6.625
Inequality in the attain	132	0	0.1960	0.0121	0.1387	0.0130	0.0615
IDH - 2013	132	0	0.6894	0.0136	0.1567	0.3370	0.5700
1. Governança	132	0	53.38	1.88	21.59	15.22	37.11
EPI Score	132	0	52.48	1.44	16.55	18.43	39.30
EV - Water Resources	132	0	27.77	2.85	32.71	0.00	0.54
EV - Agriculture	132	0	71.37	1.65	18.98	14.66	58.52
EV - Forests	132	0	39.72	2.51	28.87	0.00	18.04
EV - Fisheries	132	0	18.089	0.964	11.080	0.000	15.057
Happy Planet Index	132	0	42.673	0.778	8.936	22.591	36.587
GINNI Index	132	0	39.024	0.705	8.097	24.820	33.065
Democracy Index - 2014	132	0	5.873	0.180	2.068	1.490	4.100

1.4. ANÁLISE COMPARATIVA E CONSIDERAÇÕES

Aqui são apresentados os gráficos gerados pela funcionalidade *Graphical Summary* de acordo com os passos:

Primeiro: clique em *Stat* em seguida clique em *Basic Statistics*, e em *Graphical Summary*.

Segundo: Aparecerá uma janela e deverá ser escolhida a coluna (variável) com a qual deseja construir o gráfico. Não é necessário alterar o grau de confiança de 95,0. Clique OK e o gráfico será gerado em uma janela. Copie e cole no seu trabalho.

4.1 Definições e Análise das Variáveis deste estudo

4.1.1 Biodiversity and Habitat

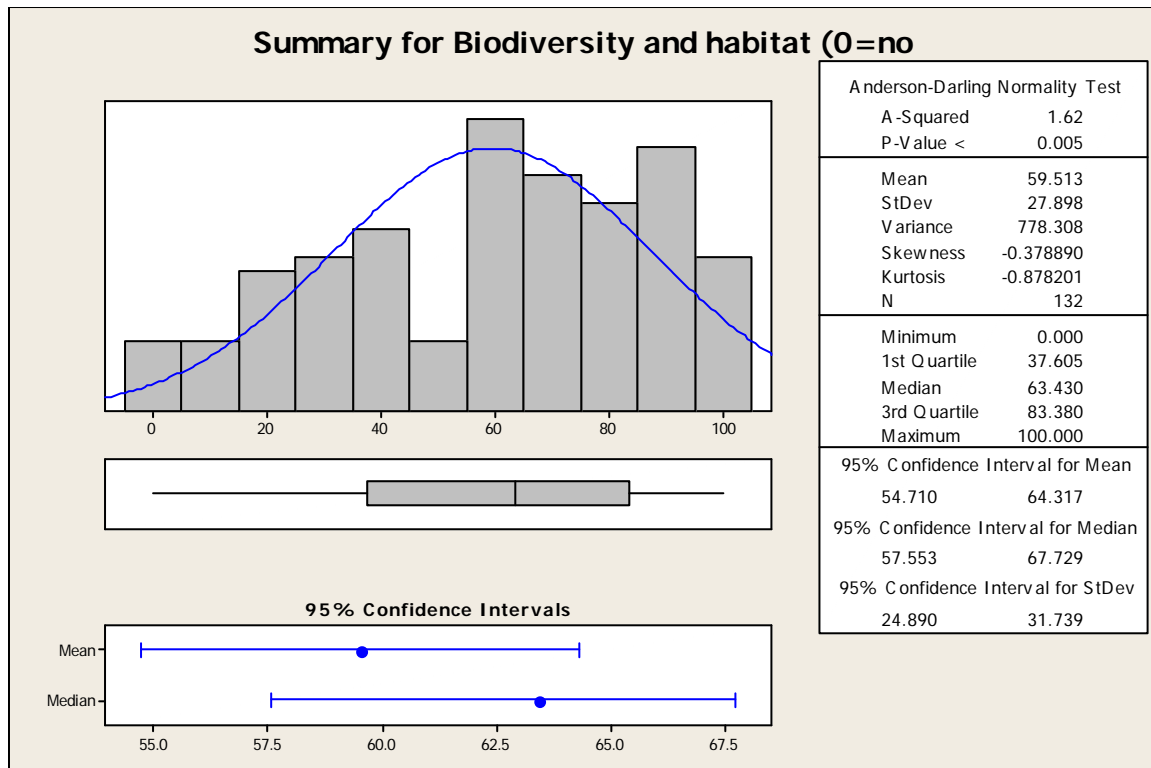


Figure 03: graph - Biodiversity and habitat (0=no protection; 100=high protection)

Definição:

A variável *Biodiversity and Habitat* (Biodiversidade: Capital Natural e Social), é formada por componentes que retratam o nível de proteção das áreas terrestres e marinhas, bem como as espécies ameaçadas ou em via de extinção e seus biomas.

source: http://epi.yale.edu/files/biodiversity_and_habitat.xls (tradução livre do autor).

Análise da variável:

A variável Biodiversity and Habitat (Biodiversidade: Capital Natural e Social) se apresenta numa escala de 0 a 100, onde o valor 0 representa nenhuma proteção à biodiversidade, e 100, alta proteção.

A curva de densidade no histograma acima indica uma distribuição próxima da normal entre os dados analisados, porém levemente deslocada para a direita. O que é evidenciada pelos valores da média (59.513) e a mediana (63.430), que resultaram relativamente próximas em seus valores. Pelo fato da média e mediana figurarem a direita do gráfico, isso nos mostra que temos mais países ativamente empenhados em cuidar de seu bioma, que o contrário. Por isso vemos 49 países abaixo da média (59.513) em contraste com outros 73 acima da média (59.513). Somente 11 países estão na faixa do valor máximo na escala, o que representa menos de 10% do total. Entre os quais somente Alemanha, Estônia, Eslovênia, Suíça, Zâmbia, Botswana, apresentaram valor máximo (100=*high protection*). O resultado do Brasil foi de 66.74 pontos, o que o coloca acima da média, e no estrato onde figuram 17 países, tais como Bulgária, Dinamarca e Egito, para citar alguns.

Identificamos 5 países no estrato com os valores mais baixos do gráfico (<5), são eles, Bósnia e Herzegovina, Iraque, Líbano, Lesoto e Libéria. O Iraque é o único com valor 0 para esta variável (0=no protection). Este país também figura na lista de países com os menores índices

de Governança e de Bem Estar para sua população e aparece no estrato dos maiores emissores de poluentes na atmosfera (Greenhouse gas emissions > 2000).

Com 95% de confiança, podemos afirmar que a média encontra-se no intervalo entre 54.710 e 64.317 pontos.

4.1.2 SPI – Social Progress Index – Índice de Progresso Social

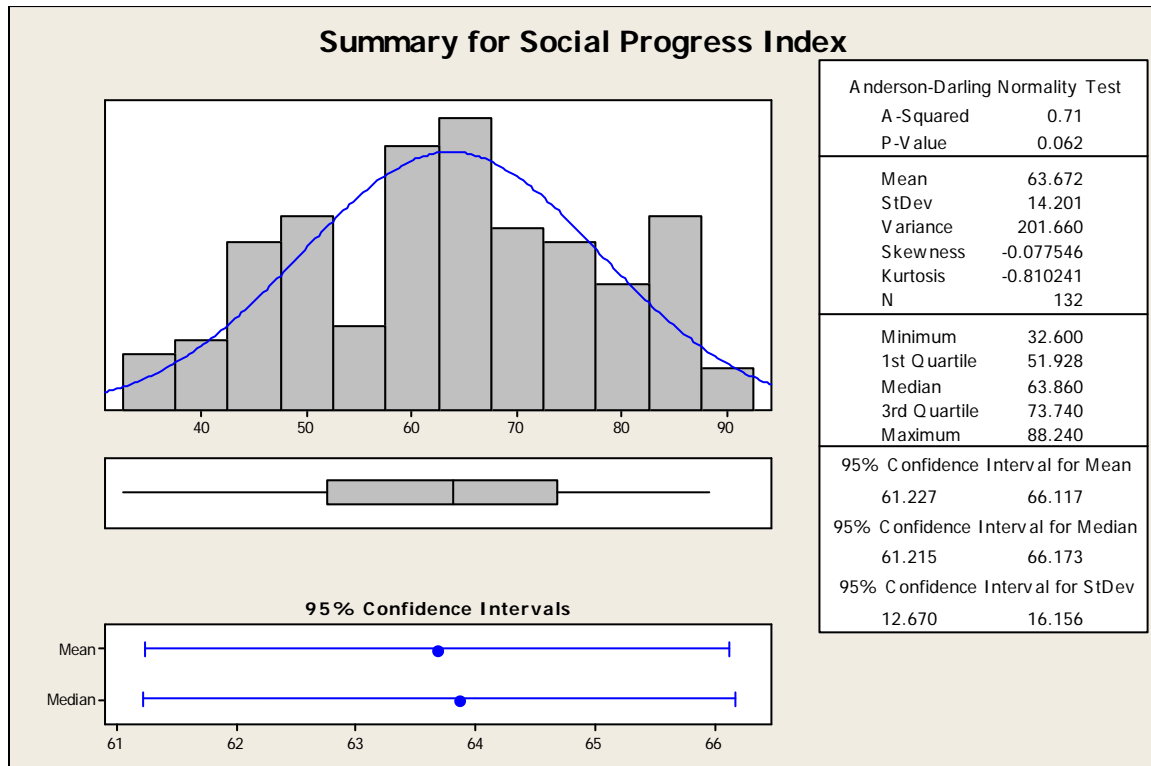


Figure 04: graph - Social Progress Index

Definição:

O Índice de Progresso Social (*Social Progress Index*) é o resultado de um processo de dois anos de pesquisas envolvendo uma equipe renomada de especialistas que inclui, entre outros, os economistas Hernando de Soto e Michael Porter e da presidente da Fundação Rockefeller, Dr. Judith Rodin. O índice sintetiza uma extensa quantidade de dados pesquisados para identificar as dimensões do desempenho das sociedades e medir o progresso social de forma abrangente e rigorosa. O índice foi estruturado em torno de 12 componentes e 54 indicadores distintos consolidados em três dimensões do Progresso Social: Necessidades Humanas Básicas, Fundações de Bem-estar e Oportunidade. A primeira dimensão, necessidades humanas básicas, por exemplo, avalia o quão bem um país prevê as necessidades essenciais de seu povo através da medição que considera se as pessoas têm comida suficiente e estão recebendo cuidados médicos básicos, se têm acesso à água potável, acesso a uma habitação condigna com serviços básicos, e se eles estão seguros e protegidos.

Fonte: origem e definição dados pelo portal: <http://pt.knoema.com/jklbfre/the-social-progress-index> (tradução livre do autor).

Análise da variável:

Considerando os 132 países analisados, o histograma acima apresenta uma distribuição próxima da normal.

Há uma concentração maior de países no estrato entre 62.5 e 67.5, totalizando 21 países. A média (63.672) e a mediana (63.860) são praticamente idênticas, diferenciando-se por 0.2 (dois décimos) apenas. Somente 3 países ocupam a posição mais elevada do gráfico, compreendida entre 87.5 e 92.5, que são os países Suíça, Nova Zelândia e Islândia. Ao passo que no estrato de 32.5 e 37.5, que representa os menores valores obtidos nesta análise, há 4 observações, representadas pelos países Burundi, República da África Central, Chade e Guiné. Cerca de 70 países estão acima da média (63.672) do índice (Upper Middle Social Progress Countries).

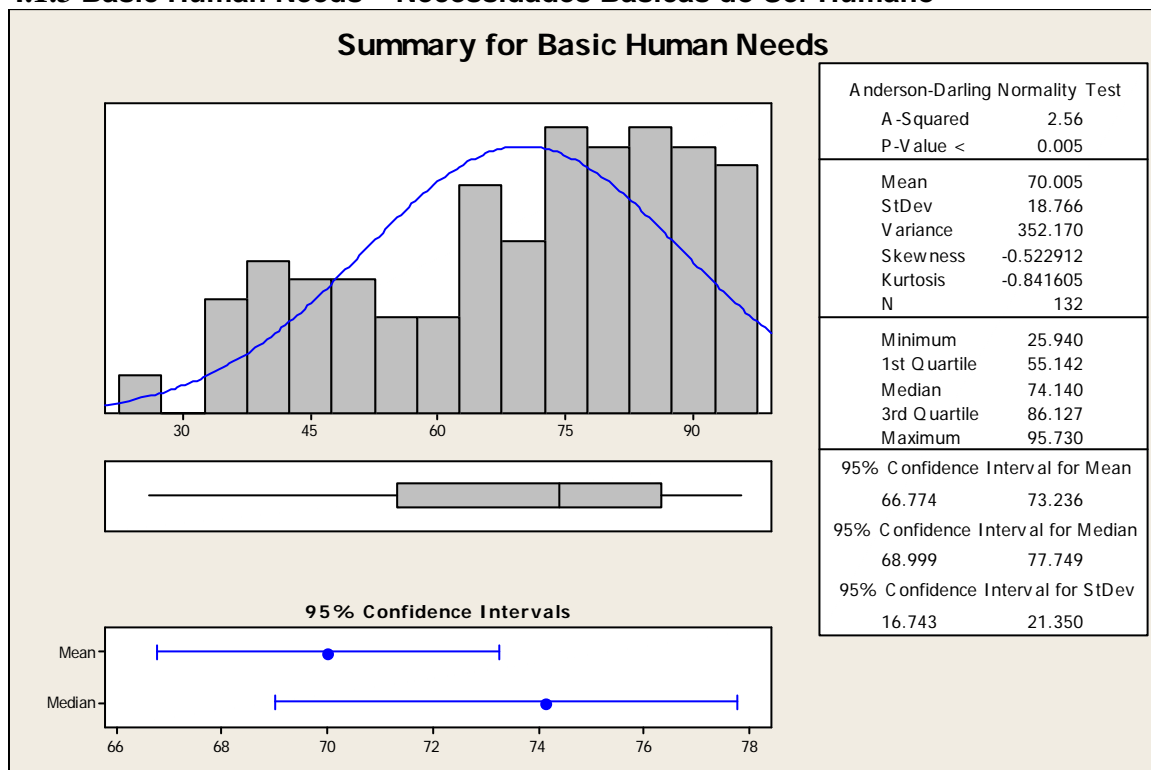
4.1.3 Basic Human Needs – Necessidades Básicas do Ser Humano

Figure 05: graph - Basic Human Needs

Definição:

Necessidades básicas do ser humano (alimentação básica e cuidados médicos)

Análise da variável:

Pela distribuição dos dados no histograma, vemos que esta distribuição não é uma normal. A média (70.005) e a mediana (74.140) apresentam valores relativamente altos, corroborando com o que nos mostra o gráfico acima, 73 países acima da média. Não houve observações no estrato entre 27.5 a 32.5

4.1.4 Foundations of Wellbeing - Fundamentos do Bem-Estar

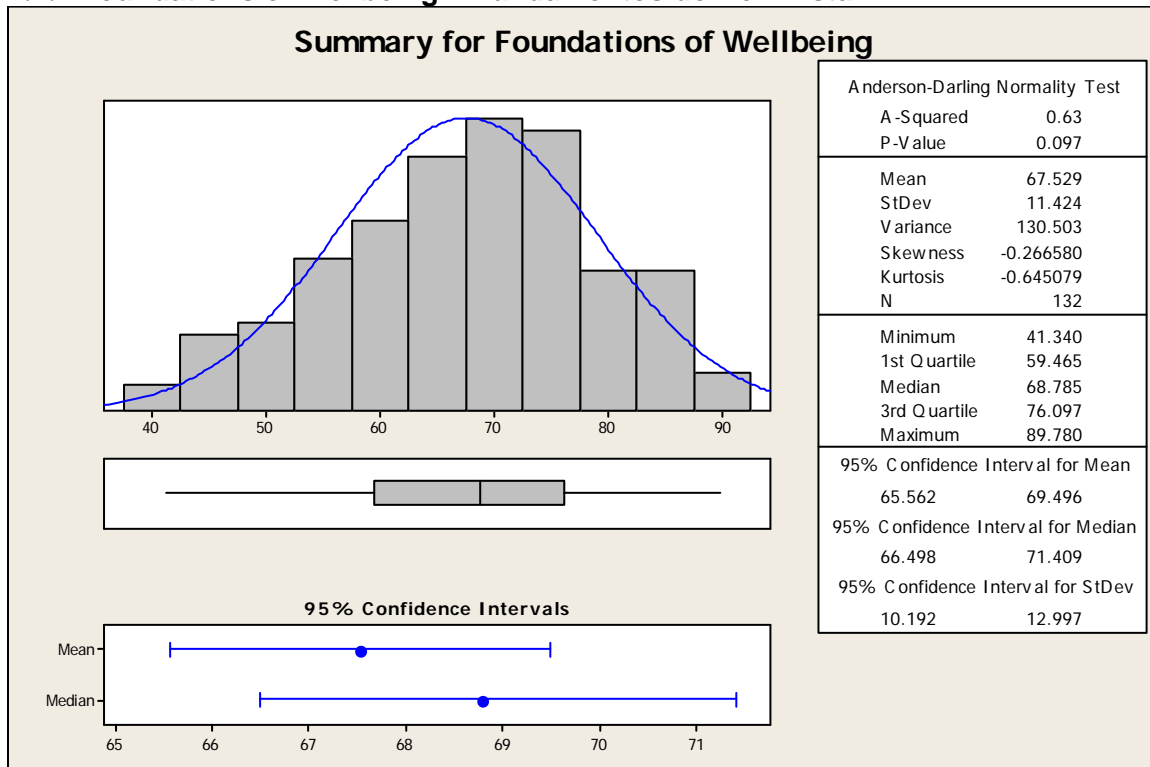


Figure 06: graph - Foundations of Wellbeing

Análise da variável:

O gráfico dos Fundamentos do Bem-Estar e melhoria de vida através da educação, mostra que a distribuição aproxima-se da normal, considerando os 132 países. A média está com o valor de 67.529, a mediana com 68.785. O estrato 67.5 e 72.5 representa a maior concentração das observações, com 23 países.

4.1.5 Opportunity

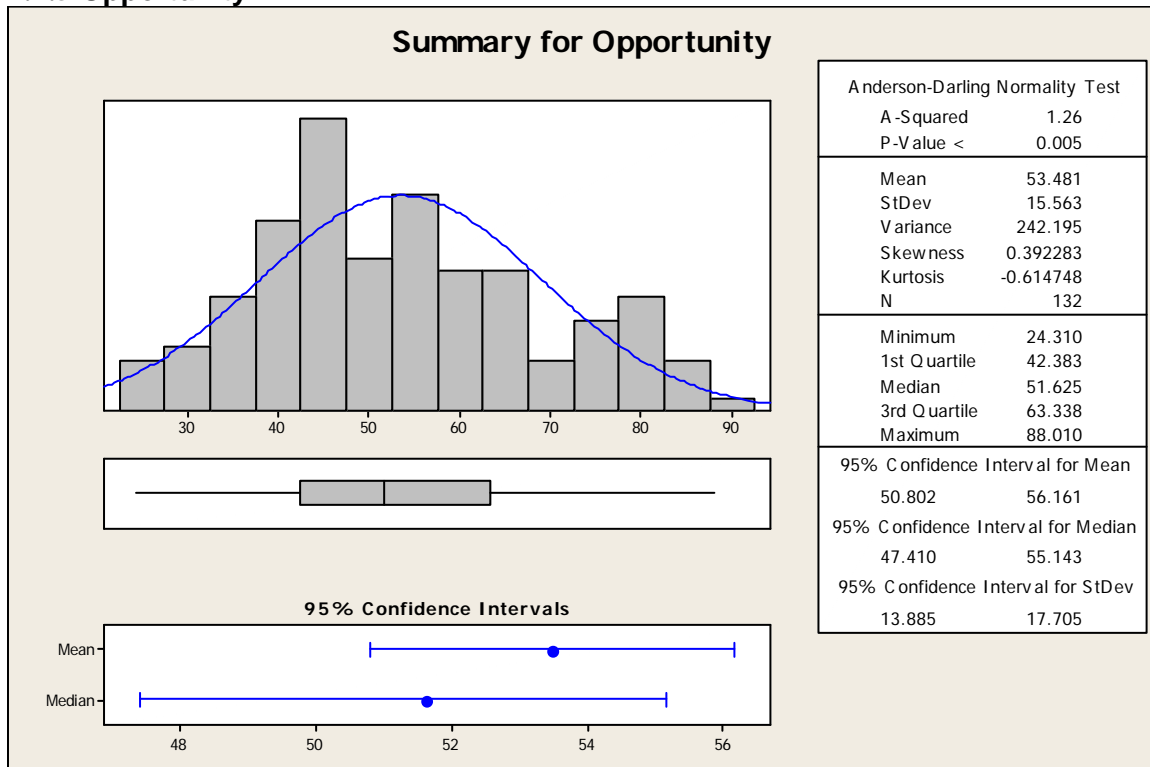


Figure 07: graph - Opportunity

Análise da variável:

Este indicador é composto por: Direitos individuais, Liberdades individuais, Tolerância e inclusão e Acesso à educação superior. E busca mostrar o quanto as pessoas são livres para exercerem seus direitos. O histograma mostra uma distribuição próxima da normal e concentra um total de 95 países no estrato entre 37.5 e 67.5. Apenas 1 país figura no estrato mais elevado do gráfico cujos valores estão entre 87.5 e 92.5, este país é a Nova Zelândia. A maior concentração de observações se encontra no estrato de 42.5 a 47.5, com 23 países.

4.1.6 Health and Wellness – Saúde e Bem-Estar

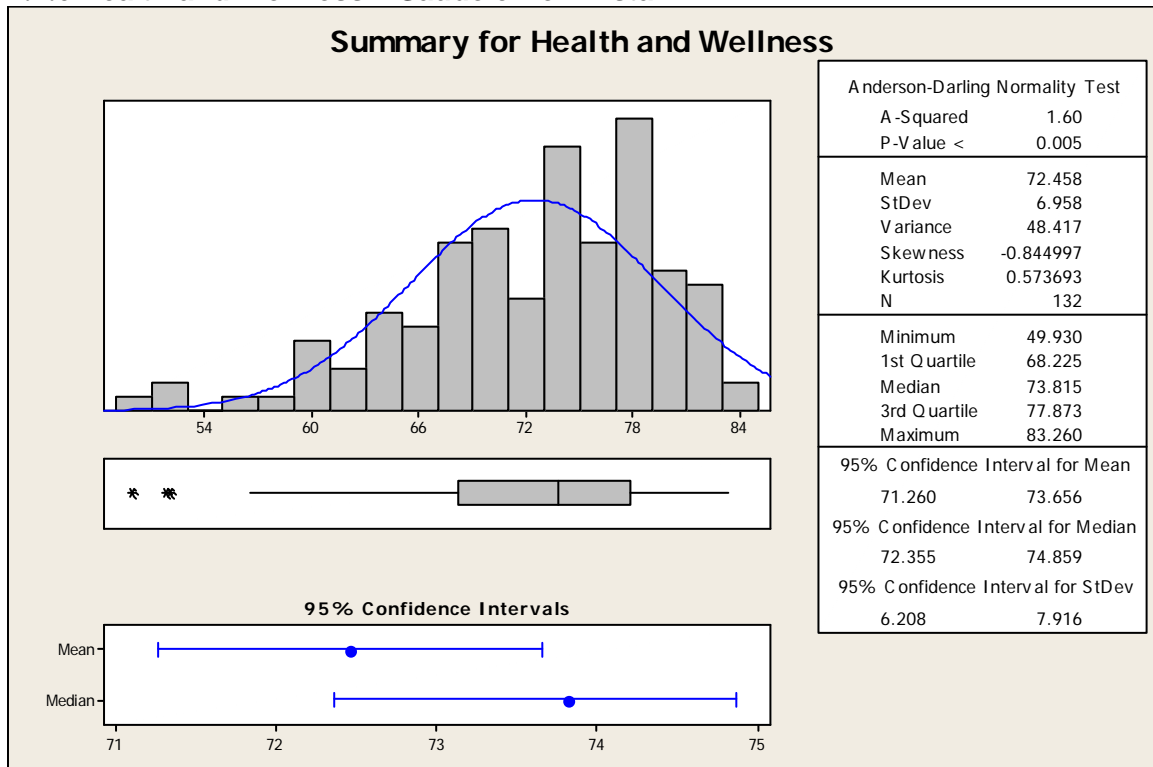


Figure 07: graph - Health and Wellness

Análise da variável:

Este indicador busca evidenciar se as pessoas vivem vidas longas e saudáveis em seus países. O histograma mostra uma distribuição próxima da normal, porém deslocada para a direita do gráfico, onde se concentra a maior parte dos dados, com média de 72.458 e mediana com valor de 73.815. O terceiro quartil com valor de 77.873, é representado pelo estrato com 21 países.

A análise nos revela 3 observações apontadas como *outliers* na parte esquerda do histograma, são eles: Cazaquistão, Rússia e Ucrânia.

4.1.7 Personal Freedom and Choice - Liberdade na tomada de decisão

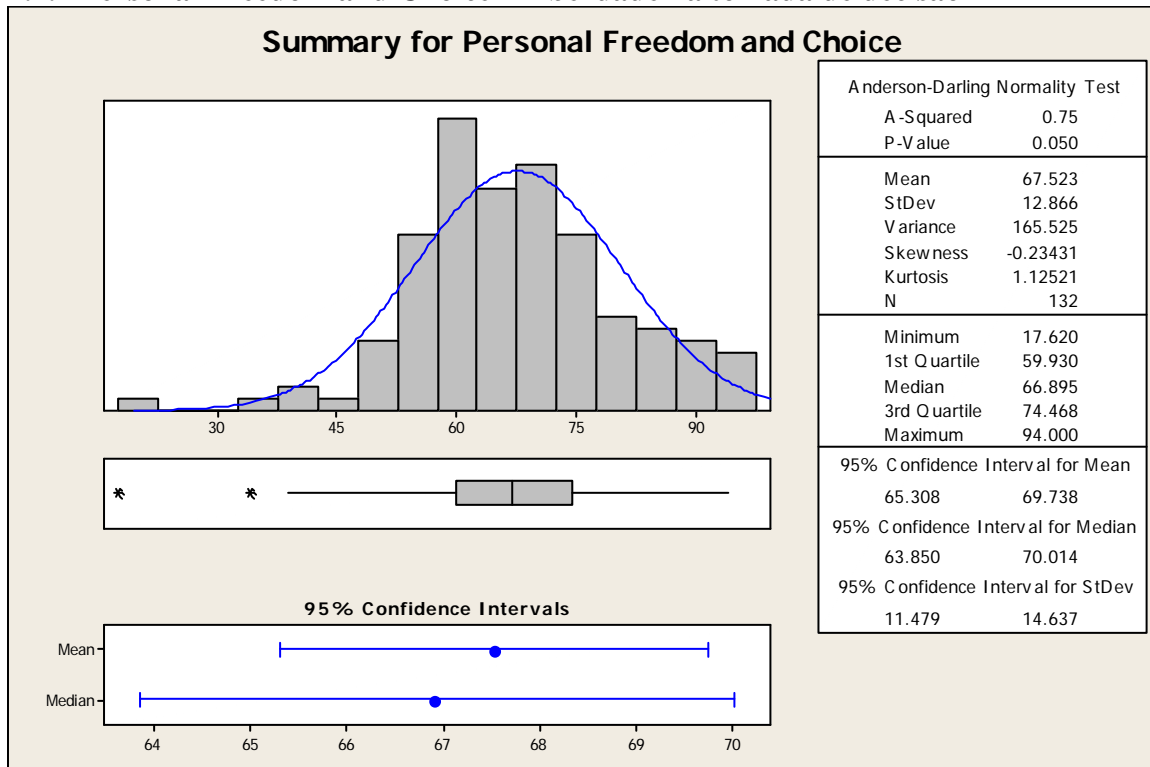


Figure 08: graph - Personal Freedom and Choice

Análise da variável:

Esta variável visa medir a liberdade na tomada de decisão e escolhas de vida do indivíduo e da população de uma país.

A distribuição aproxima-se da normal. Percebe-se uma concentração dos valores na parte direita do gráfico. A média tem valor 67.523 e a mediana com 66.895

A análise revela a presença de 2 *outliers* no conjunto de 132 países analisados, são eles: Sudão (34.05) e Mauritânia (17.62). A maior concentração de observações, se encontra no estrato entre 52.5 e 77.5, com 95 países.

4.1.8 Greenhouse gas emissions - Emissões de gases de efeito estufa

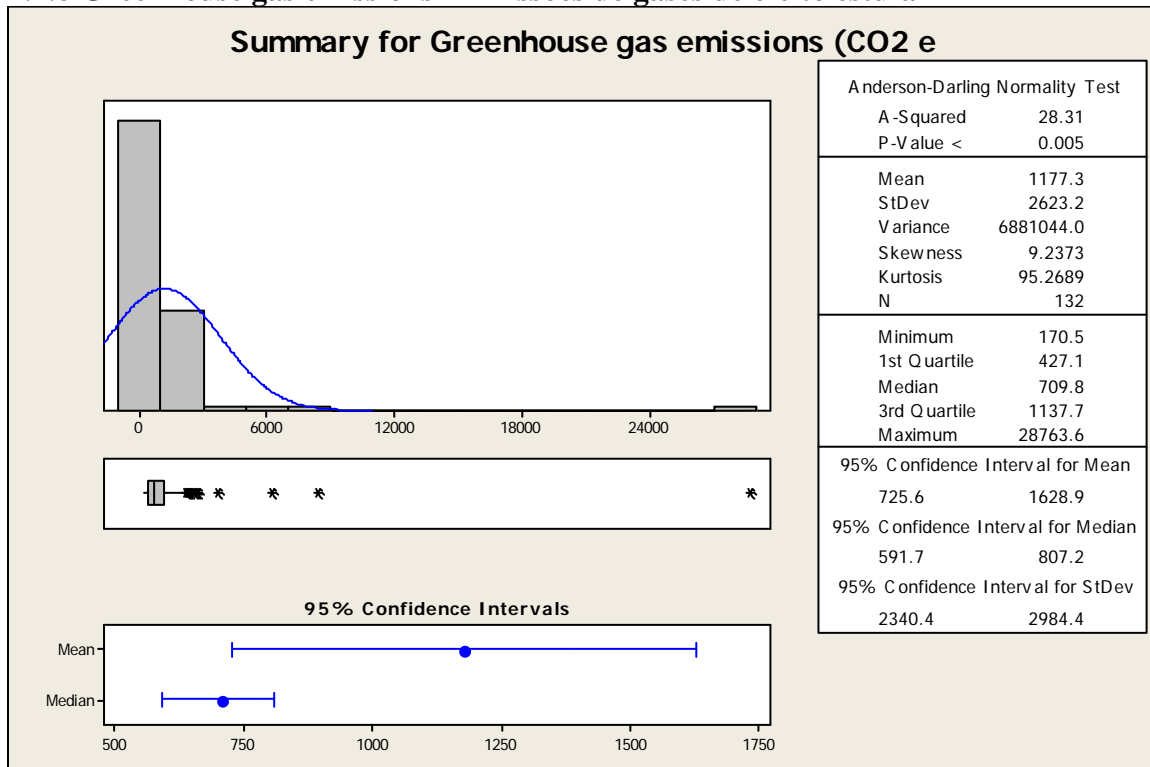


Figure 09: graph - Greenhouse gas emissions (CO₂ equivalents per GDP)

Análise da variável:

As emissões de dióxido de carbono (CO₂), metano (CH₄), óxido nitroso (N₂O), hidrofluorcarbonos (HFCs), perfluorcarbonos (PFCs) e hexafluoreto de enxofre (SF₆), expressas em equivalentes de CO₂ usando 100 anos potenciais de aquecimento global encontrados no Painel Intergovernamental de Mundaça de Clima - Segundo Relatório de Avaliação por PIB-PPP. No modelo SPI, os dados são escalados 0-4: 4: <100 3: 100-200 2: 200-1000 1: 1000-2000 0:> 2000. Este indicador captura a quantidade de emissão de poluentes na atmosfera.

A média apresenta valor de 1177.3 e a mediana de 709.8. Dos 132 países observados, 95 deles estão no estrato entre -1000 e 1000. Observa-se a presença de 4 *outliers* nesta análise, sendo: Burundi (8442.23), República Africana Central (28763.64), Montenegro (6220.0) e Zâmbia (3734.91).

4.1.9 Water withdrawals as a percentage of resources - Uso de água como porcentagem dos recursos

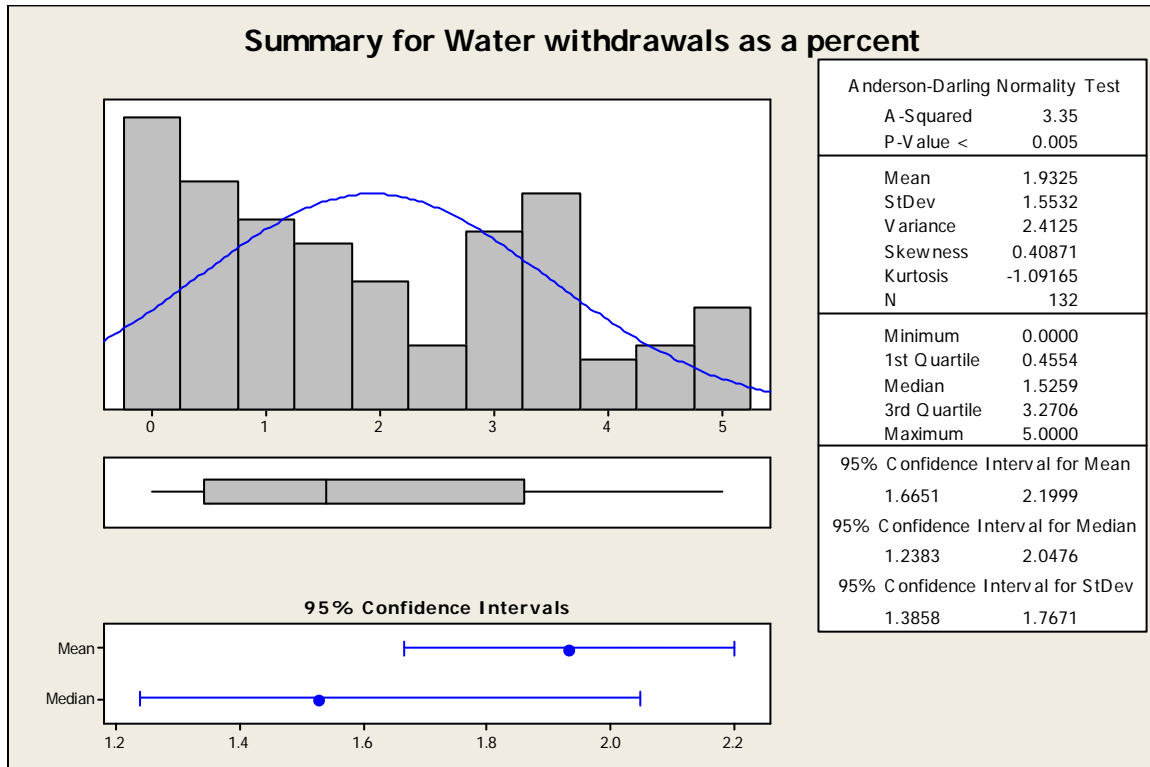


Figure 10: graph - Water withdrawals as a percentage of resources

Análise da variável:

Esta variável demonstra o uso de água como porcentagem dos recursos naturais. E apresenta uma escala de 0 a 5, sendo: De 4–5: Extremamente alto (>80%), de 3–4: Alto (40–80%), de 2–3: Médio-Alto (20–40%), de 1–2: Médio-Baixo (10–20%) e de 0–1: Baixo uso (<10%). A curva de distribuição aproxima-se da normal. E 41 países estão abaixo do coeficiente 1, denotando que aproximadamente um terço dos países analisados mostram eficiência na gestão do recurso. Enquanto 8 países figuram na faixa “Extremamente alto (>80%)”, o maior valor atingido nesta escala é 5, onde figuram 3 países, como segue: Jamaica, Trinidad e Tobago e União dos Emirados Árabes. A média é de 1.9325 e a mediana tem valor de 1.5259.

4.1.10 Modern slavery, human trafficking and child marriage - privação da liberdade e trabalho forçado, o tráfico de pessoas e exploração, e o casamento forçado de indivíduos menores de 18 anos

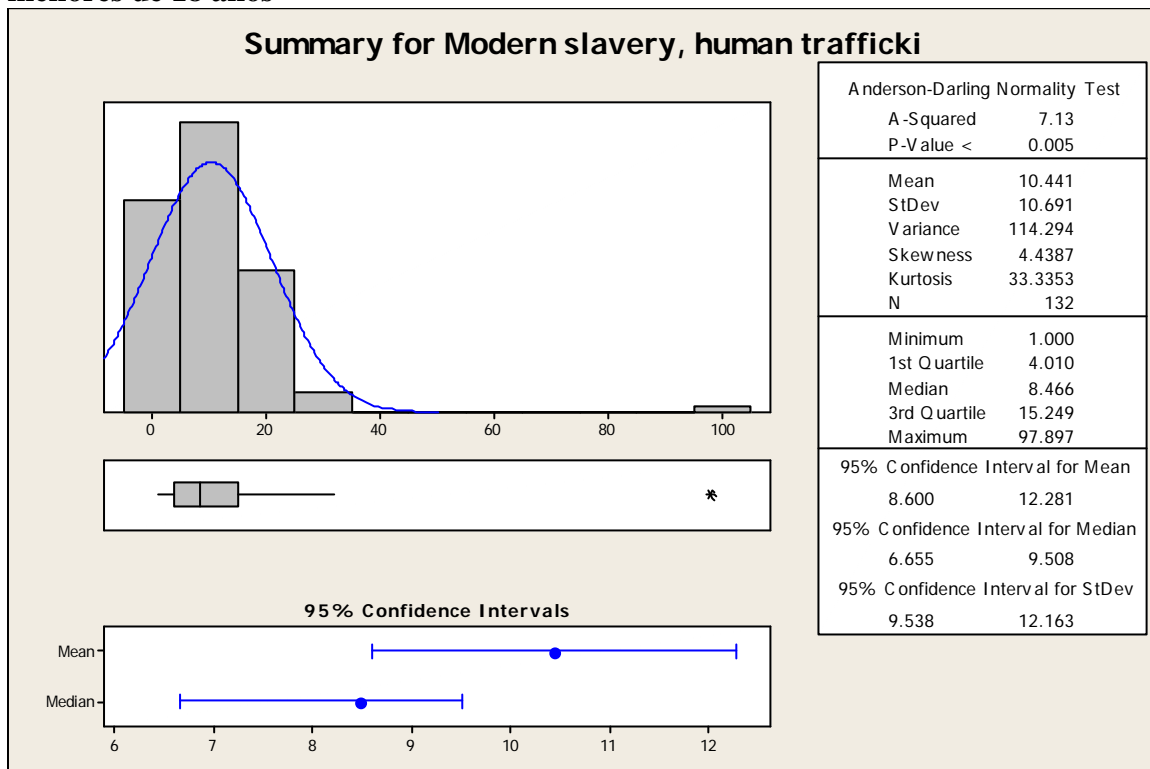


Figure 11: graph - Modern slavery, human trafficking and child marriage (1=low; 100=high)

Análise da variável:

Esta variável é a combinação de 3 variáveis que medem: a escravidão nos tempos atuais (privação da liberdade e trabalho forçado), o tráfico de pessoas e exploração, e o casamento forçado de indivíduos menores de 18 anos.

Os dados em sua maioria, total de 127 observações, se concentram na parte esquerda do gráfico. Há a presença de 1 *outlier*, com valor de 97.90 cujo país correspondente é a Mauritânia. A média é de 10.441 e a mediana de 8.466.

4.1.11 Satisfied demand for contraception (% of women) - Demanda por contracepção atendida

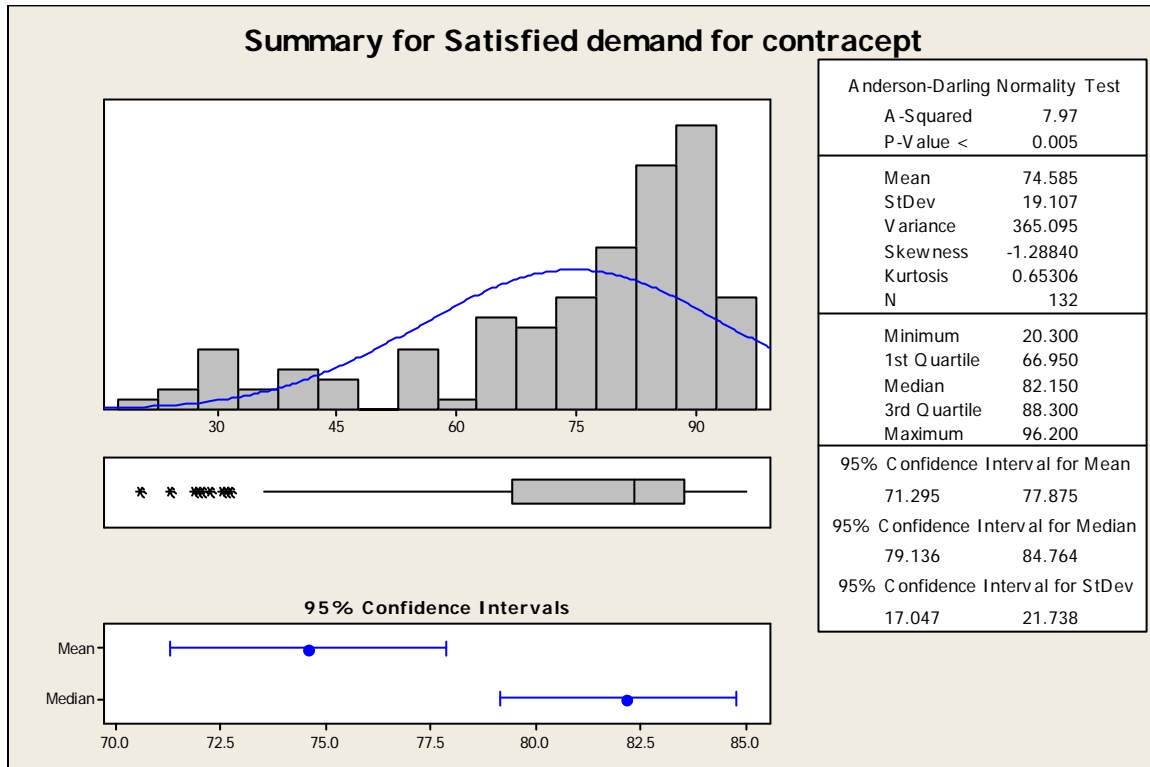


Figure 12: graph - Satisfied demand for contraception (% of women)

Análise da variável:

A variável “Demanda por contracepção atendida”, mede o percentual de demanda total de planeamento familiar entre mulheres casadas ou em união com idades entre 15 e 49 anos que estão satisfeitas com métodos contraceptivos disponíveis.

Podemos observar que 87 países figuram acima da média que tem valor de 74.585, isso indica que esta demanda está sendo bem atendida.

A mediana tem valor de 82.150 e o histograma destaca a presença de 9 países como *outliers* nas observações, são eles: Guinéa (30.50), Togo (31.60), Angola (31.20), Sudão (29.00), Mauritânia (28.10), Senegal (27.70), Libéria (27.00), Mali (24.20) e Chade (20.30).

4.1.12 Women treated with respect – Mulheres tratadas com respeito e dignidade

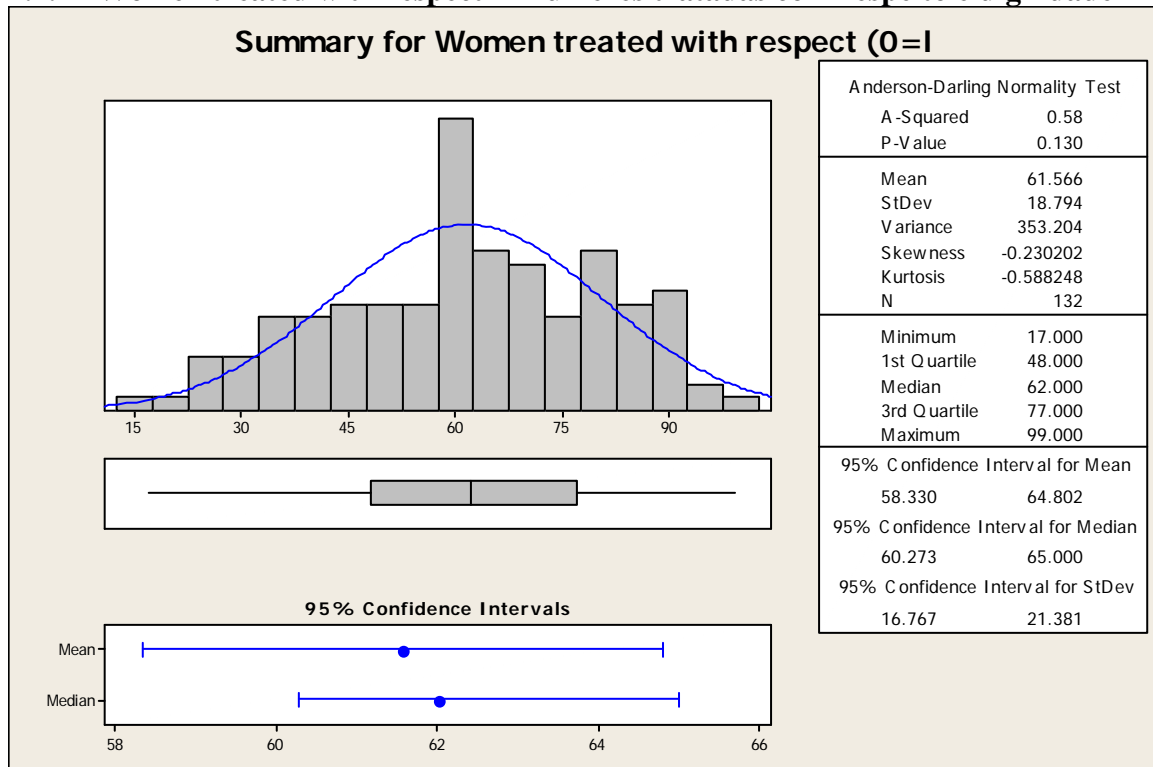


Figure 13: graph - Women treated with respect (0=low; 100=high)

Análise da variável:

Variável que contabiliza respostas SIM (positivas) de mulheres respondentes a respeito de sua percepção sobre serem tratadas com respeito e dignidade em seus países.

A curva de distribuição se aproxima de uma normal. A média tem valor de 61.566, a mediana é 62.000, onde se encontra a maior quantidade de observações agrupadas, totalizando 22 países.

4.1.13 Tolerance for immigrants – Tolerância a Imigrantes de outros países

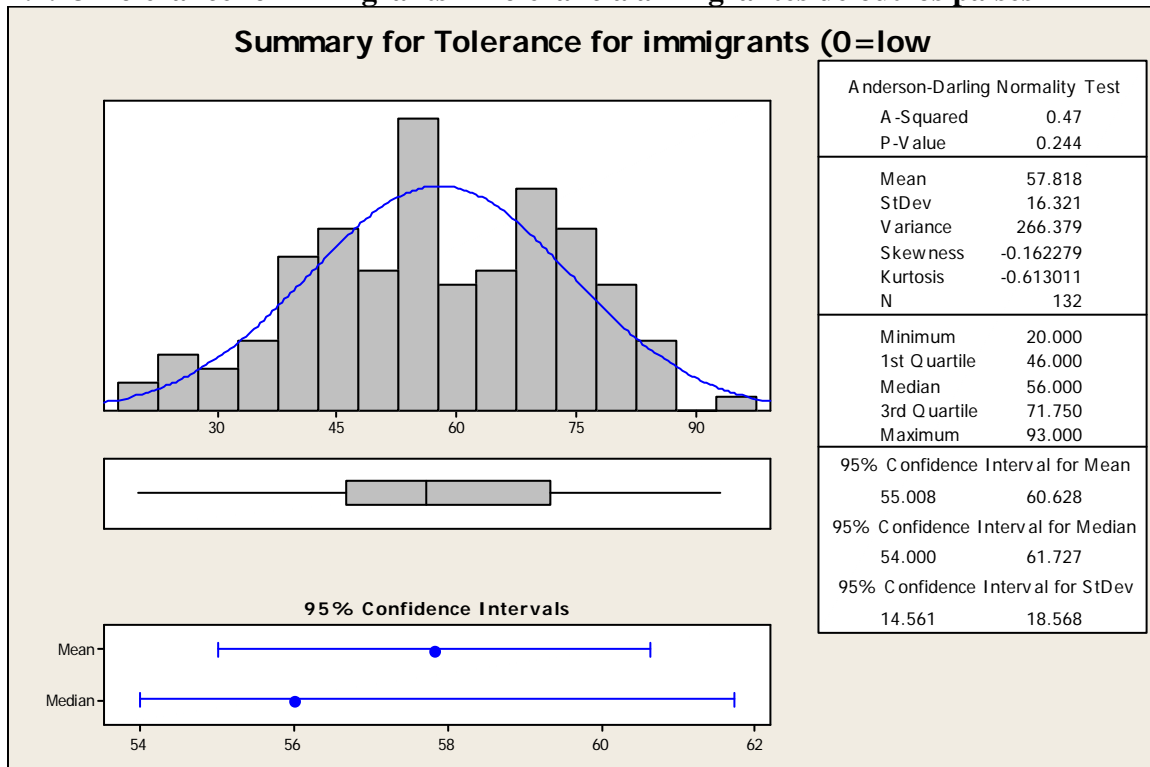


Figure 14: graph - Tolerance for immigrants (0=low; 100=high)

Análise da variável:

Tolerância a imigrantes: O percentual de entrevistados respondendo sim para a pergunta: "Será que a cidade ou área onde você vive é um bom lugar ou não é um bom lugar para viver para imigrantes de outros países?"

Variável que contabiliza respostas SIM (positivas) da população sobre a percepção de onde vivem ser um bom lugar para imigrantes de outros países.

A curva se aproxima de uma distribuição normal. A média apresenta valor de 57.818 e a mediana de 56.000, região esta do gráfico onde concentra o maior valor observado, composto por 21 países.

Nota-se a ausência de observações no estrato entre 87.5 e 92.5. E somente uma observação na barra de valor máximo do gráfico, ocupada pelo país Canadá (93.00). Em contrapartida, temos 2 observações no estrato de 17.5 a 22.5, denotando países menos tolerantes ao imigrante, são eles: Indonésia (22.00) e Malásia (20.00).

4.1.14 Tolerance for homosexuals - Tolerância a homossexuais

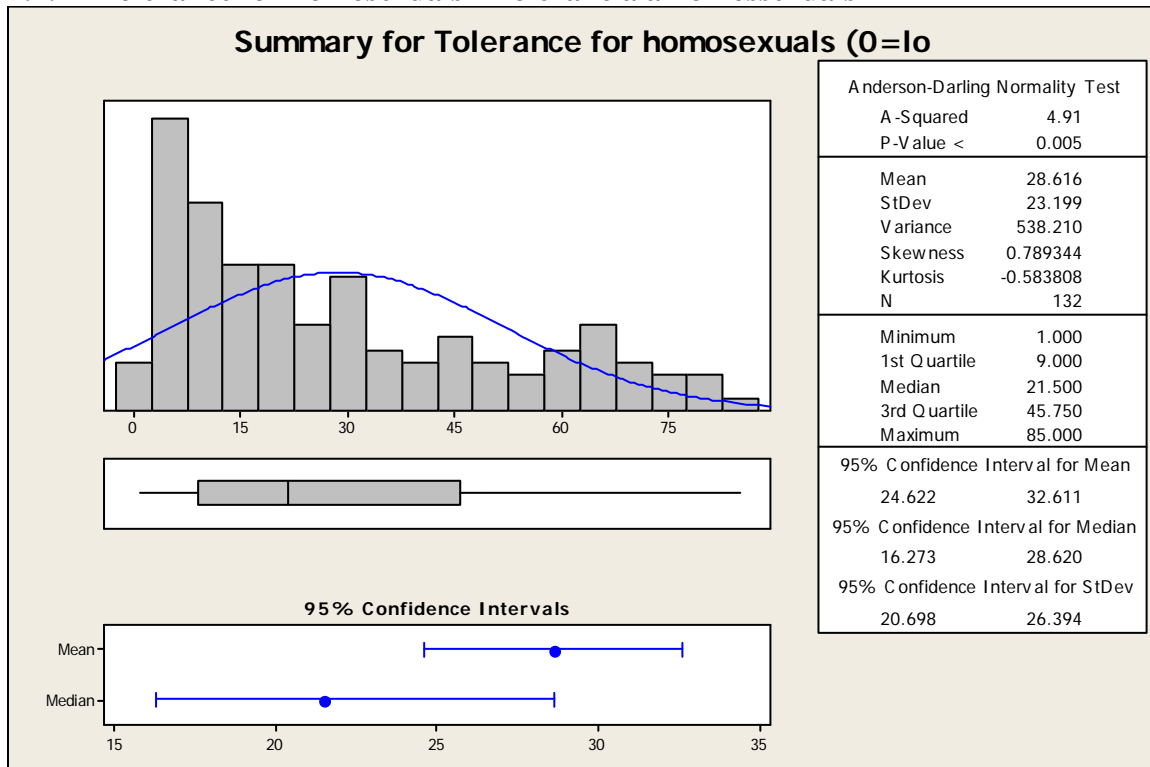


Figure 15: graph - Tolerance for homosexuals (0=low; 100=high)

Análise da variável:

Tolerância a homossexuais: O percentual de entrevistados que respondem sim à pergunta: "Será que a cidade ou área onde você vive é um bom lugar ou não é um bom lugar para se viver sendo gays ou lésbicas?"

Esta variável que contabiliza respostas SIM (positivas) da população sobre a percepção de onde vivem ser um bom lugar para a comunidade LGBT (gays, lésbicas,...).

A média tem valor de 28.616 e a mediana de 21.500, o que é mostrado no gráfico com a maior quantidade de observações figurarem no lado esquerdo do gráfico, cujo maior valor está no estrato entre 2.5 e 7.5, com 24 países que demonstram baixa tolerância. Do outro lado, no estrato de 82.5 e 87.5, com apenas um país, figura a Holanda (85.00) com o maior índice de tolerância encontrada nas observações.

4.1.15 Discrimination and violence against minorities - Discriminação e violência contra minorias

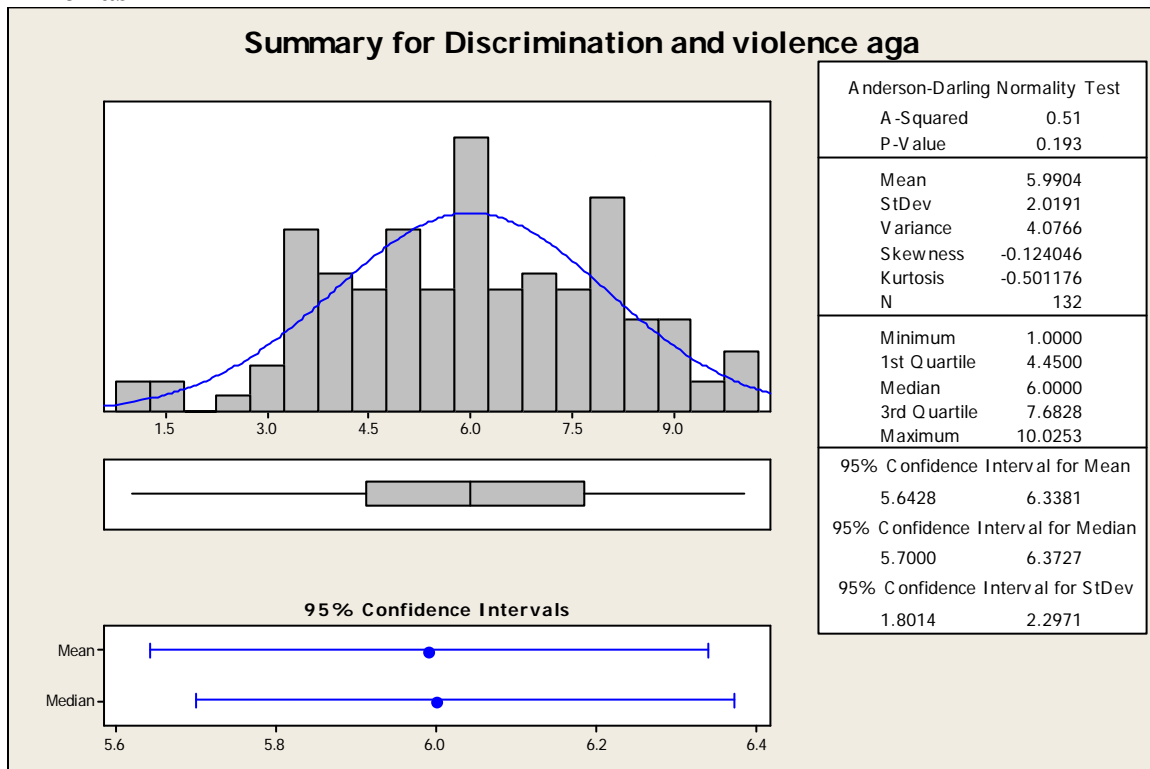


Figure 16: graph - Discrimination and violence against minorities (0=low; 10=high)

Análise da variável:

Discriminação e violência contra minorias: Indicador de Reclamações em Grupo. A discriminação, falta de poder, violência étnica, a violência comunitária, a violência sectária e violência religiosa, medido em uma escala de 0 (baixa pressão) a 10 (pressões muito altas). A variável mede o grau de intolerância, discriminação e violência contra as minorias e desfavorecidos (éticas, religiosas,...).

A curva se aproxima de uma distribuição normal. A média tem valor de 5.9904 e a mediana de 6.0000, onde se encontra o maior agrupamento de países, totalizando 18 observações. Não há dados no estrato entre 1.75 e 2.25, portanto há 4 países no estrato entre 0.75 e 1.75, são eles: Suécia (1.00), Irlanda (1.60), Islândia (1.00) e Finlândia (1.40), representando o grupo que apresenta o menor indicador de discriminação e violência contra minorias. E na outra extremidade na faixa entre 9.75 e 10.25, países como Iraque (10.03), Israel (9.75), Nigéria (9.80) e Sudão (9.96), apresentam os maiores índices.

4.1.16 Religious tolerance - Tolerância religiosa

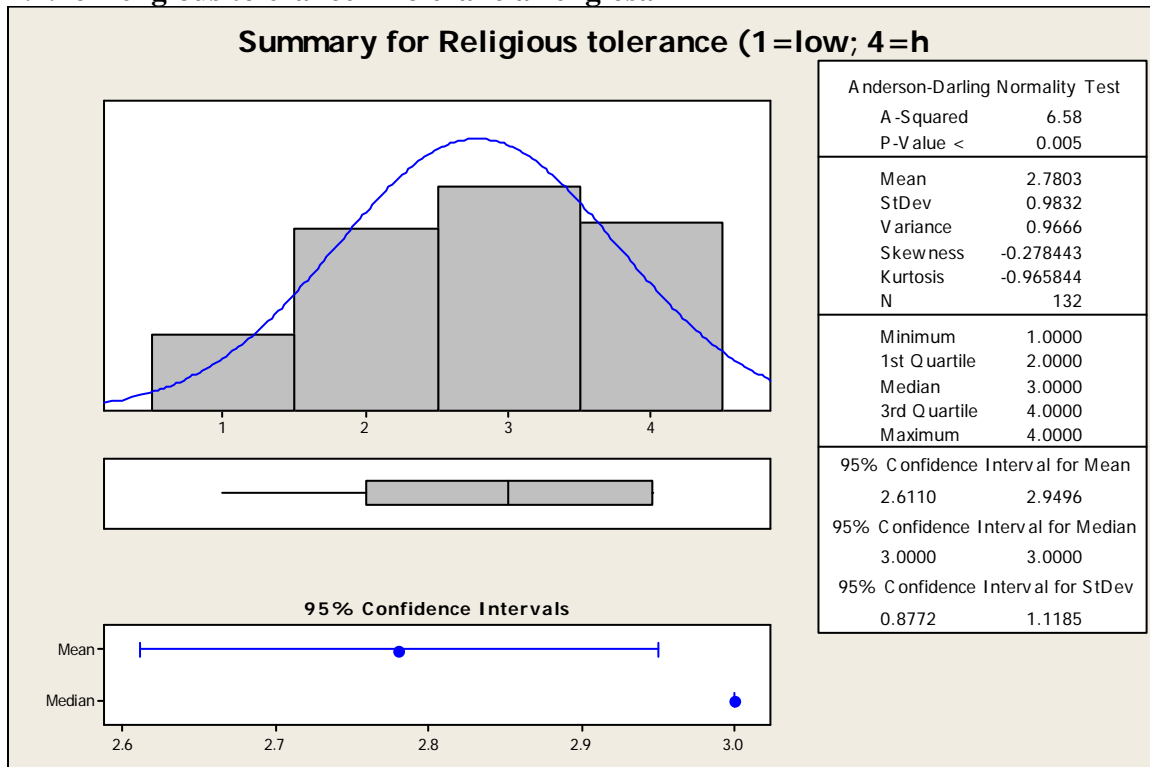


Figure 17: graph - Religious tolerance (1=low; 4=very high)

Análise da variável:

Tolerância religiosa: Uma medida de 13 tipos de hostilidade religiosa por indivíduos, organizações ou grupos da sociedade, incluindo conflitos relacionados com religião armados ou terrorismo, multidão ou a violência sectária, o assédio sobre traje ou por razões religiosas ou outras intimação ou abuso relacionadas com a religião. No modelo SPI, os pontos variam de 1 (baixo) a 4 (muito alto).

Esta variável é a medida de 13 tipos de hostilidades causadas por indivíduos, organizações ou sociedade (assédio, abuso, intimidação,..)

A curva de distribuição se aproxima de uma normal. A média tem valor de 2.7803 e a mediana de 3.0000. O terceiro quartil tem valor 4.0000, o mesmo do valor máximo.

Das observações, notamos que há 81 países nas faixas que representam tolerância “alta” e “muito alta”, ao passo que 15 países estão na faixa de “baixa” tolerância às expressões religiosas.

4.1.17 Community safety net - Apoio familiar e comunitário

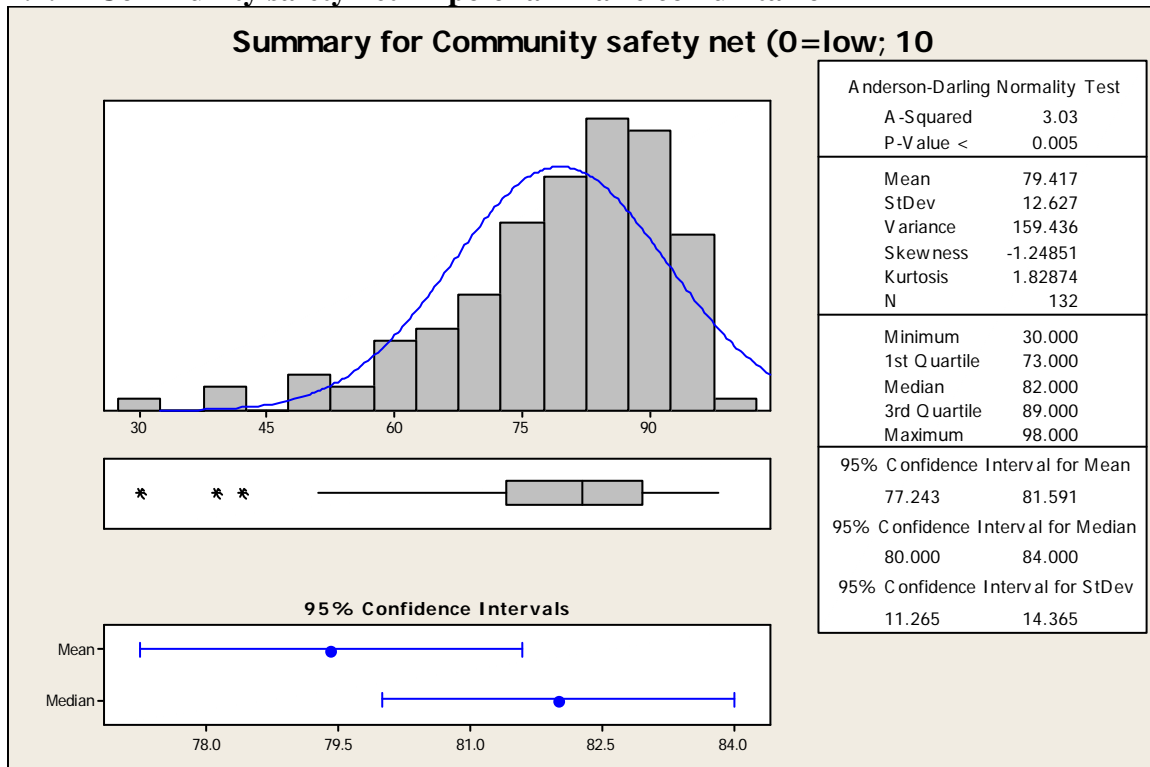


Figure 18: graph - Community safety net (0=low; 100=high)

Análise da variável:

Apoio familiar e comunitário: O percentual de entrevistados que respondem sim à pergunta: "se você estivesse em apuros, você tem parentes ou amigos que você pode contar para ajudá-lo sempre que você precisar deles, ou não?"

Variável que contabiliza respostas SIM (positivas) da população sobre uma situação problemática, onde possam contar com seus familiares ou amigos onde estiverem.

A curva aproxima-se de uma normal. A média tem valor de 79.417 e a mediana de 82.000. A maior concentração de observações está a direita no histograma. Dos 132 países analisados, 100 deles se encontram no estrato entre 72.5 e 97.5.

Não há observações na faixa entre 42.5 e 47.5, no entanto aparecem 3 *outliers* situados abaixo do valor 42.5, são eles: Togo (30.00), República da África Central (39.00) e Burundi (42.00), representando os países com menor índice de apoio familiar e comunitário.

4.1.18 Years of tertiary schooling - Anos de escolaridade superior

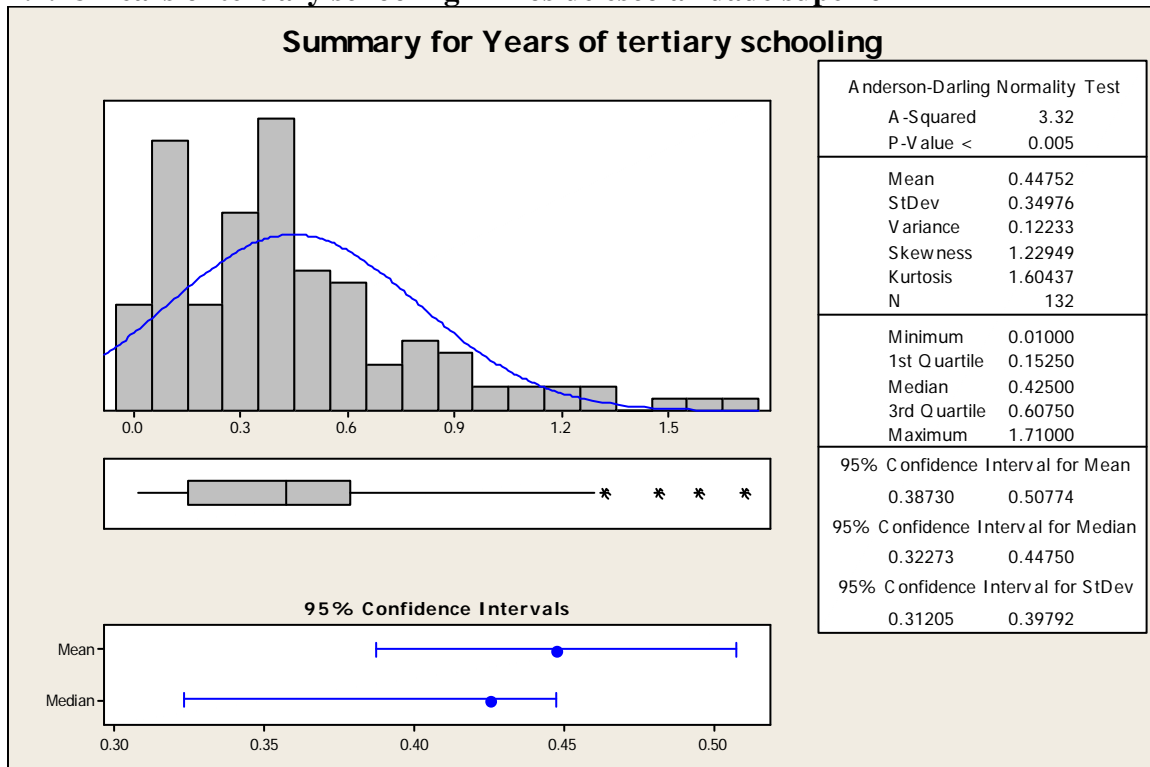


Figure 19: graph - Years of tertiary schooling

Análise da variável:

Anos de escolaridade superior

Esta variável retrata a média de anos de ensino superior concluído entre pessoas com mais de 25 anos de idade.

Nota-se a maior concentração de observações do lado esquerdo do gráfico. O que resulta em 106 países na faixa abaixo de 0.65. A média tem valor 0.44752, a mediana é 0.42500 e o valor máximo é de 1.71000. Há 4 *outliers* neste gráfico, representados pelos países Canadá (1.32), Nova Zelândia (1.47), Rússia (1.58) e USA (1.71) representando o grupo de países com maior quantidade de anos de escolaridade superior.

4.1.19 Women's average years in school - Taxa média de escolarização das mulheres

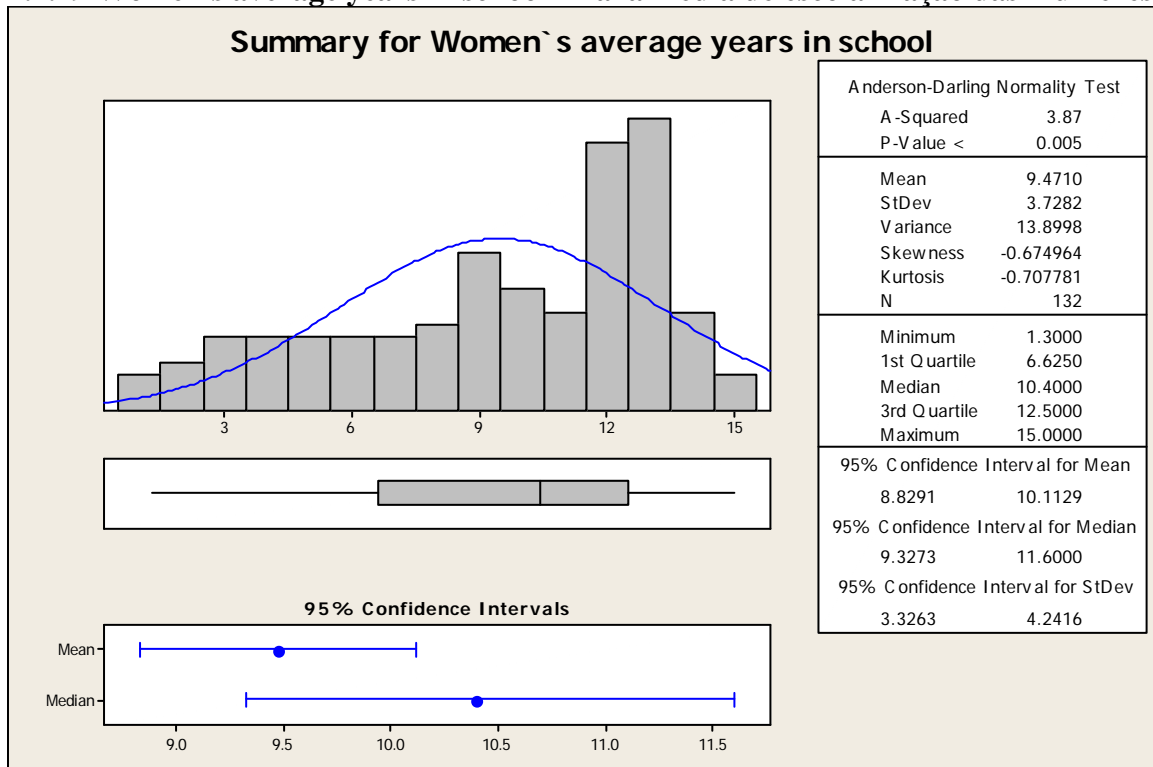


Figure 20: graph - Women's average years in school

Análise da variável:

Taxa média de escolarização das mulheres

O número médio de anos de escola frequentada por mulheres entre 25 e 34 anos de idade, incluindo fundamental, médio e ensino superior.

A curva se aproxima de uma normal. A média tem valor 9.4710 e a mediana de 10.4000

Dos 132 países estudados, vemos 48 países no estrato entre 11.5 e 13.5.

4.1.20 Inequality in the attainment of education - Desigualdade no desempenho educacional

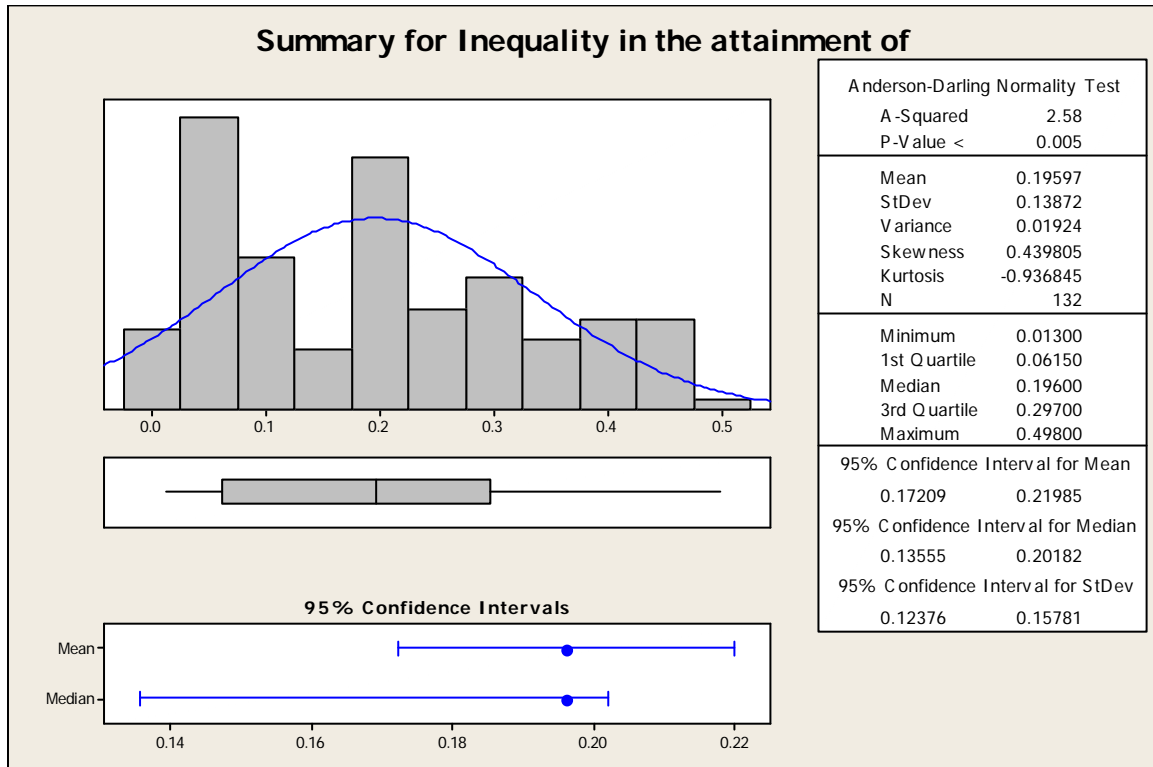


Figure 21: graph - Inequality in the attainment of education (0=low; 1=high)

Análise da variável:

Desigualdade no desempenho educacional

A perda na educação potencial devido à desigualdade, calculada como a diferença percentual entre o Índice de Educação do Índice de Desenvolvimento Humano, que compreende a média de anos de escolaridade e anos de escolaridade esperados, e o Índice de Educação Ajustado à Desigualdade.

A curva se aproxima de uma distribuição normal. A média é de 0.19597, a mediana tem valor de 0.19600, região do gráfico onde se encontra a segunda maior concentração de observações, com 25 países. Menor apenas que o estrato entre 0.025 e 0.075, que apresenta 29 países nesta faixa.

4.1.21 Índice de Desenvolvimento Humano - IDH – 2013

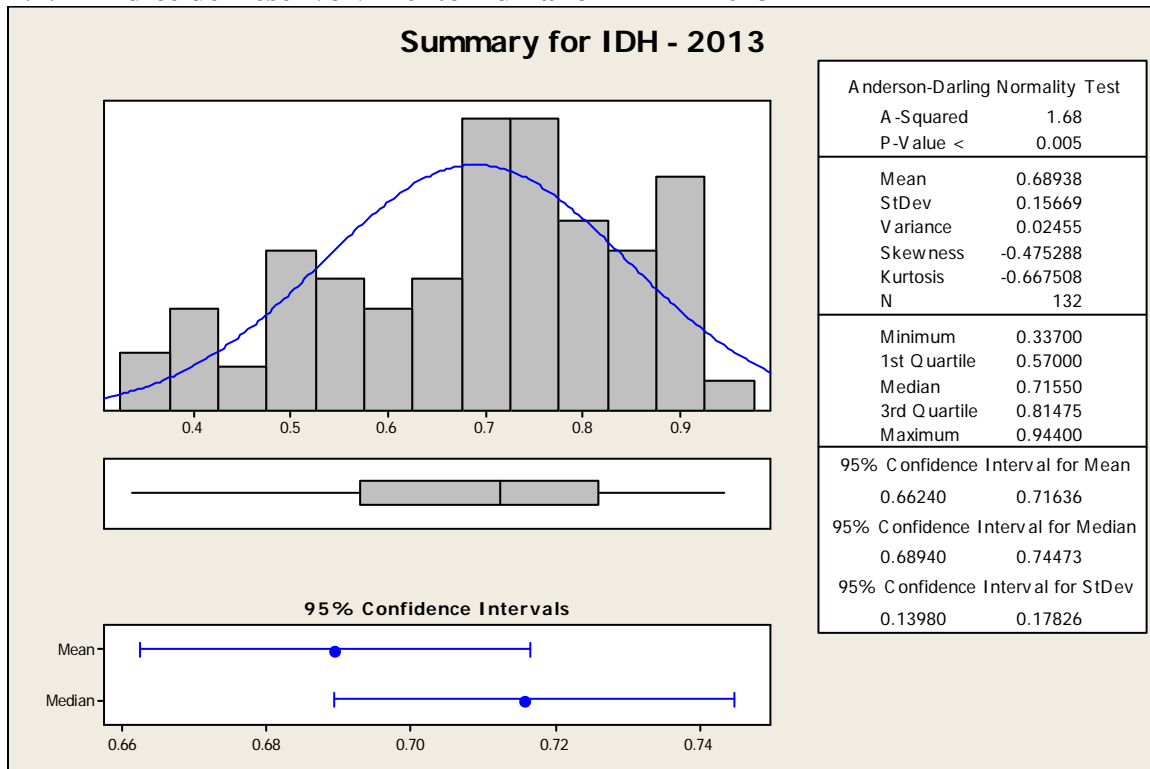


Figure 22: graph - Índice de Desenvolvimento Humano - IDH – 2013

Análise da variável:

IDH – 2013: O Índice de Desenvolvimento Humano (IDH) mede o progresso de uma nação a partir de três dimensões: renda, saúde e educação. Fonte: PNUD, 2013.

O trabalho preliminar publicado por PNUD - Programa de las Naciones Unidas para el Desarrollo sobre o desenvolvimento humano em 2014 aponta que o IDH é progresso real da população em relação a expectativa de vida, educação, saúde, habitação, segurança e condições para um desenvolvimento sustentável.

Fonte: <http://hdr.undp.org/sites/default/files/hdr14-summary-es.pdf>

O histograma acima aproxima-se de uma distribuição normal. A média tem valor de 0.68938 e a mediana de 0.71550. A parte mais expressiva das observações se encontra mais a direita do gráfico e da média. Existem 40 países entre as faixas de 0.675 e 0.775.

Do lado esquerdo do gráfico, estão Niger, República do Congo, República da África Central (0.34) e Chade (0.37), considerados pelo índice como países de baixo desenvolvimento ou subdesenvolvidos.

4.1.22 Governança

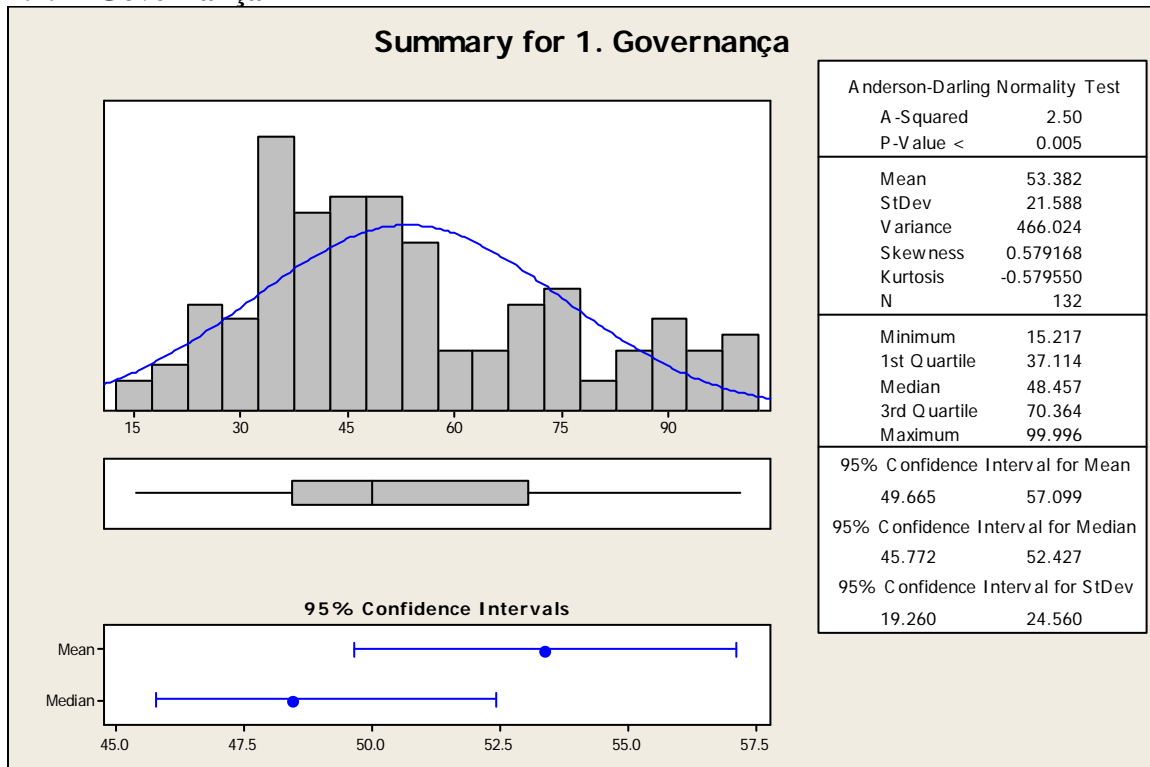


Figure 23: graph - Governança

Análise da variável:

Governança: O World Economic Fórum ressalta as estratégias de atuação: Simplificar os processos administrativos e fazer uso de alianças regionais e de tecnologia para coibir a corrupção e os abusos; Aprimorar os instrumentos e mecanismos regionais que possibilitem uma ação coletiva capaz de prevenir violações da democracia e garantir a preservação da ordem democrática (http://www3.weforum.org/docs/LA11/WEF_LA11_Report_PT.pdf).

Indicador que mede a capacidade de um país de eleger, monitorar e substituir seus governantes eleitos. Mede também nível de estabilidade política, medidas contra corrupção, entre outros.

A curva da distribuição se aproxima da normal. A média tem valor de 53.382, a mediana de 48.457 e o terceiro quartil tem valor de 70.364

A maior quantidade de observações para os 132 países analisados, está compreendida entre 32.5 e 37.5 com 18 países. Somente a Finlândia atingiu o índice 100.00 (máximo). No extremo inferior, 2 países figuram no estrato entre 12.5 e 17.5 , são eles: República da África Central (15.99) e Sudão (15.22).

4.1.23 EPI Score

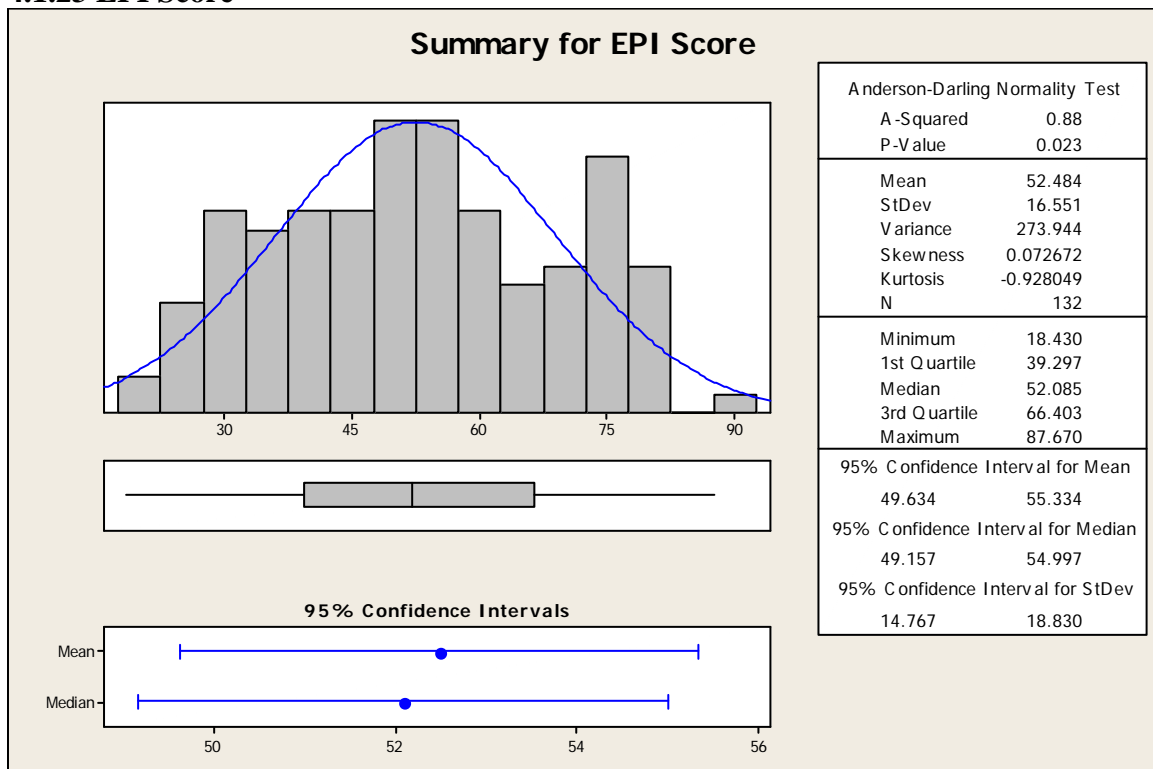


Figure 24: graph - EPI Score

Análise da variável:

EPI Score: Índice de Performance Ambiental, composto de: proteção da saúde do ser humano decorrente de ameaças ambientais e proteção do ecossistema.

A curva da distribuição se aproxima da normal, a média é de 52.484, a mediana 52.085, onde se concentram 32 países neste quartil. Os países se distribuem uniformemente no histograma. A linha se encontra praticamente no meio da caixa do box plot. Enquanto 1 país figura no limite máximo do gráfico (Suíça 87.67), 2 países se encontram no limite mínimo em relação aos 132 países analisados (Mali 18.43 e Lesotho 20.81)

4.1.24 EV - Water Resources

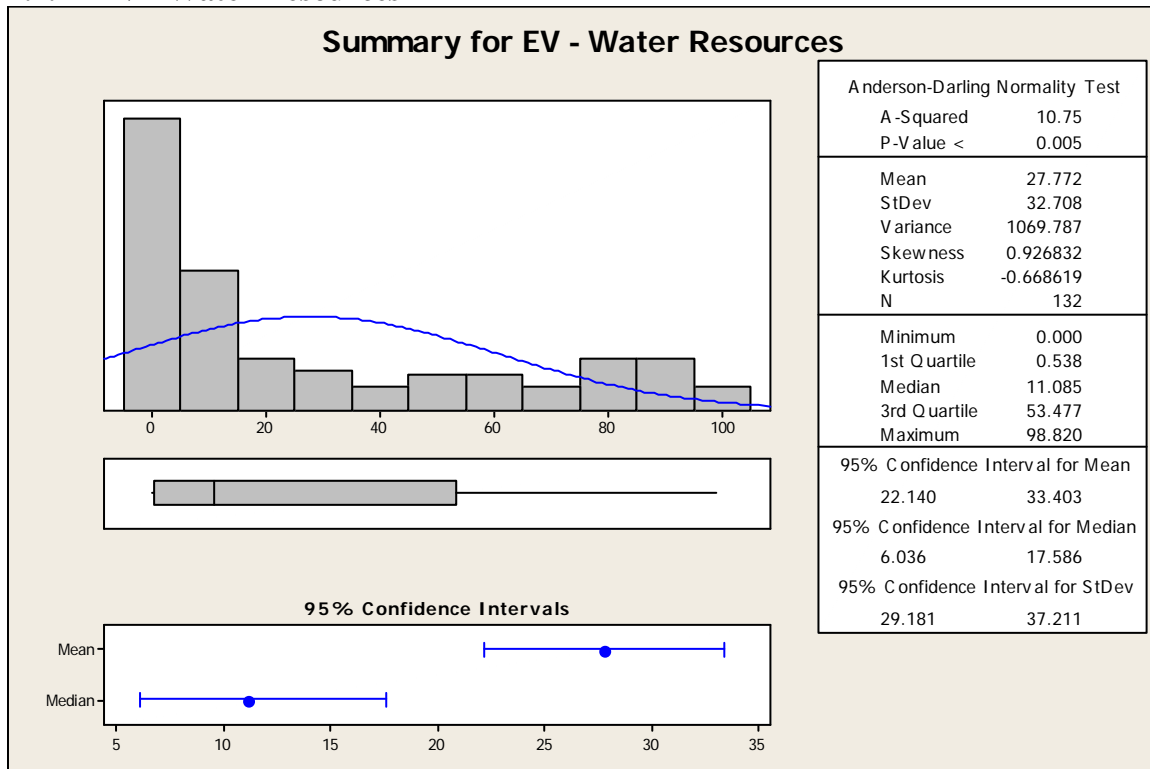


Figure 25: graph - EV - Water Resources

Análise da variável:

EV - Water Resources

Wastewater treatment level weighted by connection to wastewater treatment rate.

A curva da distribuição se aproxima da normal, a média tem valor de 27.772 e a mediana de 11.085, isso confirma a concentração das observações na parte esquerda do gráfico. Cerca de 90 observações encontram-se abaixo da média.

4.1.25 EV – Agriculture - subsídios destinados à Agricultura

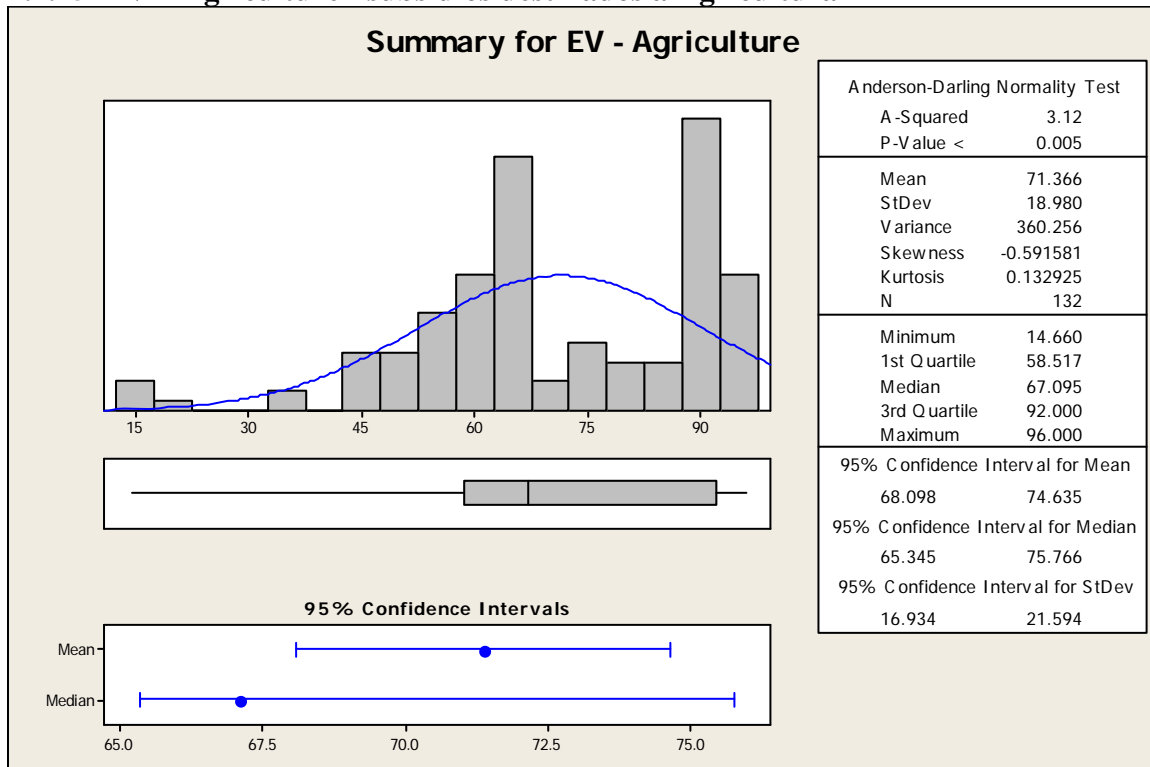


Figure 26: graph - EV – Agriculture

Análise da variável:

EV – Agriculture

Variável que mede os subsídios destinados à Agricultura e a Regulação no uso de pesticidas.

A média tem valor 71.366, a mediana de 67.095 e nota-se os dados no histograma mais concentrados a direita do gráfico.

A maior concentração de países está situada no estrato entre 87.5 e 92.5. No canto esquerdo do gráfico com baixo resultado neste índice, encontram-se Israel (14.66), Rússia (16.93), Senegal (17.45) e Botswana (20.00).

4.1.26 EV - Forests

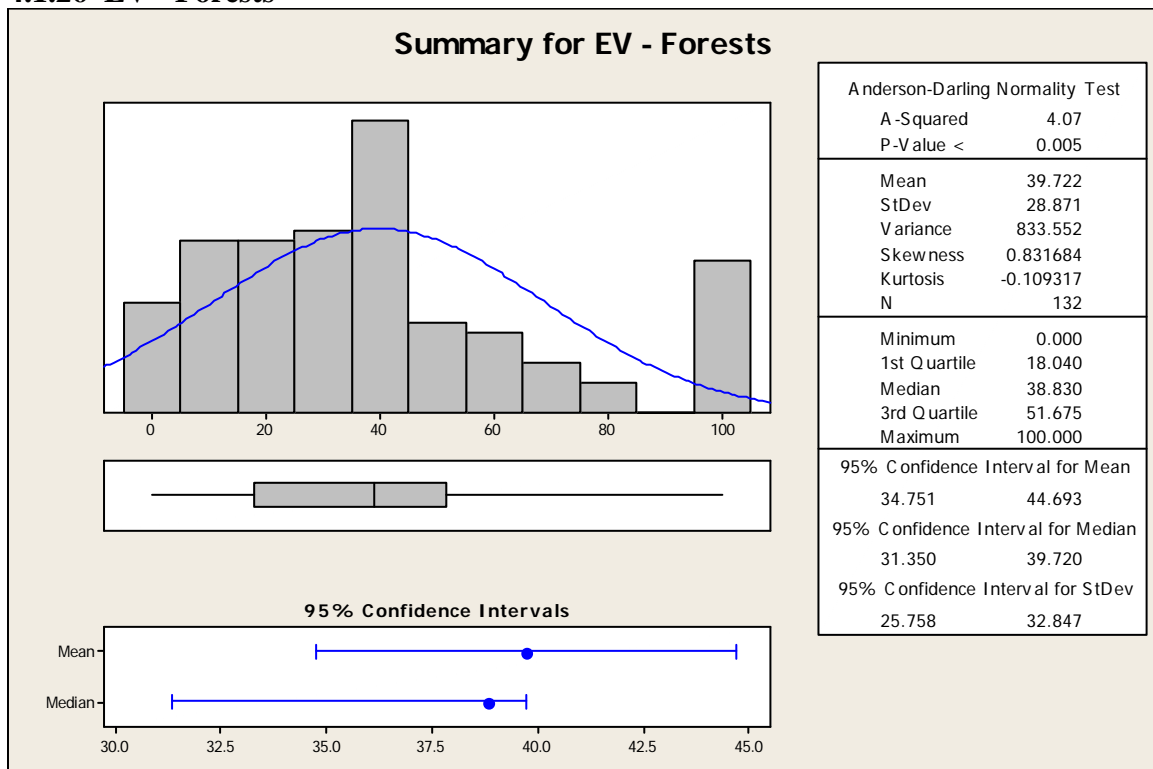


Figure 27: graph - EV - Forests

Análise da variável:

EV - Forests

Variável que demonstra a variação das áreas cobertas por floresta (desmatamento, queima e exploração).

A média tem valor de 39.722, a mediana de 38.830. Vê-se que 29 países estão localizados na faixa entre 35 e 45 pontos, a maior concentração deste gráfico. Dos 132 países analisados, somente 15 deles atingem o valor máximo no índice.

4.1.27 EV – Fisheries

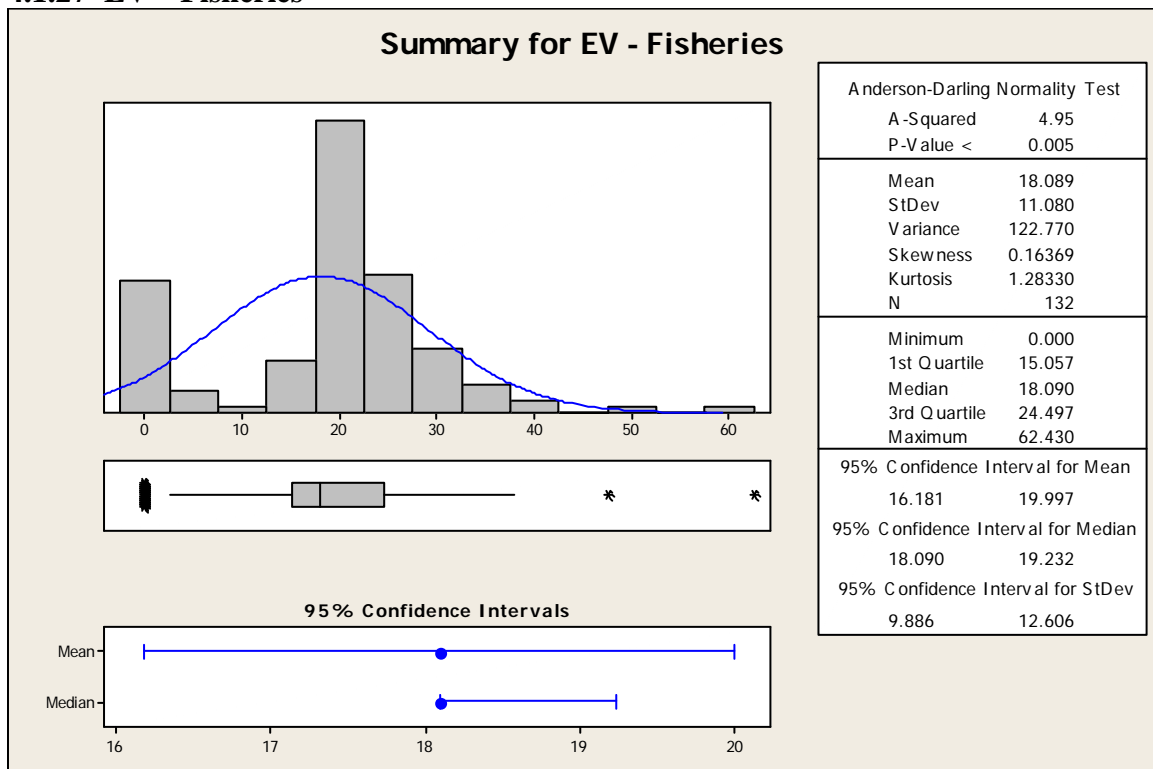


Figure 28: graph - EV – Fisheries

Análise da variável:

EV – Fisheries

Coastal Shelf Fishing Pressure and Fish Stocks

A média é 18.089, a mediana 18.090 e 51 observações no estrato entre 17.5 e 22.5.

Quanto aos *outliers*, não serão computados neste momento os casos com valor 0 (zero), pois esta revisão serão efetuadas nas próximas análises. Porém, na outra extremidade, temos 2 *outliers*, sendo Mauritius (47.50) e Sri Lanka (62.43).

4.1.28 Happy Planet Index - Bem Estar Sustentável

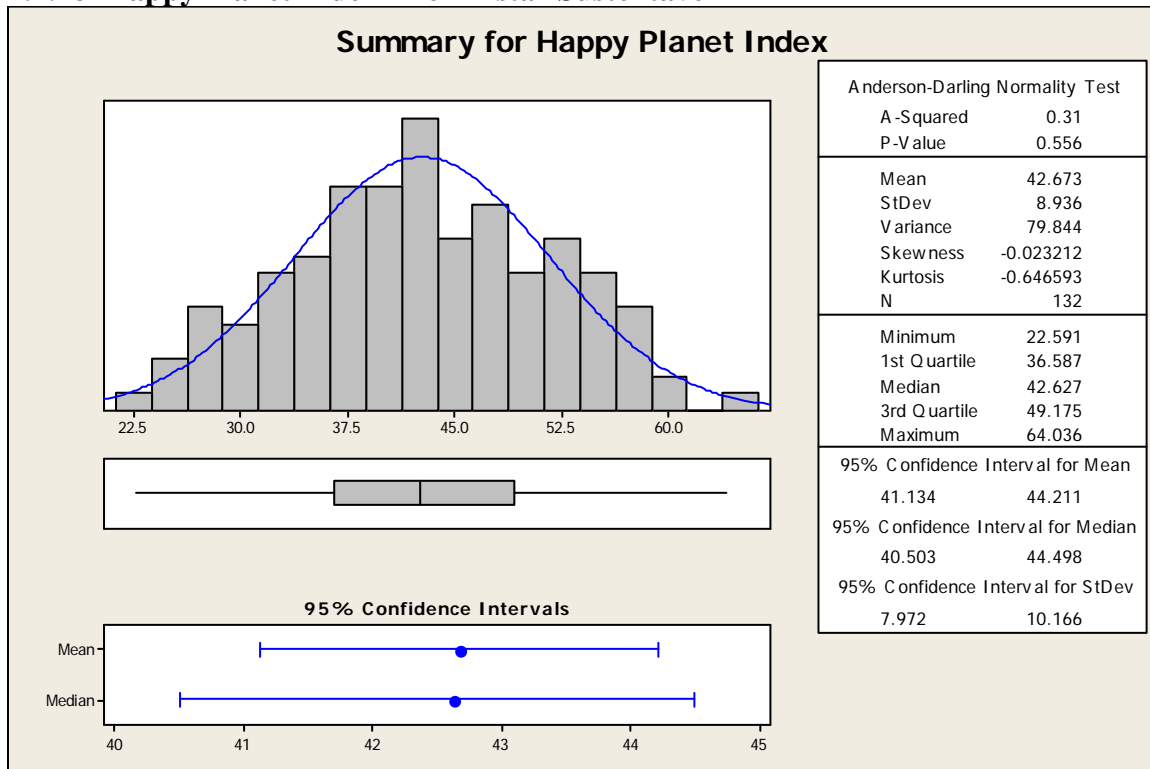


Figure 29: graph - Happy Planet Index

Análise da variável:

Happy Planet Index

HPI, medida do Bem Estar Sustentável (Longevidade, Felicidade e Vida sustentável para a população do país).

A curva da distribuição se aproxima de uma normal. A média tem valor 42.673 e a mediana com 42.627. OS dados se encontram distribuídos uniformemente seguindo a curva normal.

No estrato entre 41.25 e 43.75 encontramos 17 países. O país com o maior índice de felicidade considerando os 132 indivíduos estudados, é a Costa Rica (64.04) e com menor índice de felicidade é a Botswana (22.59).

4.1.29 GINI Index

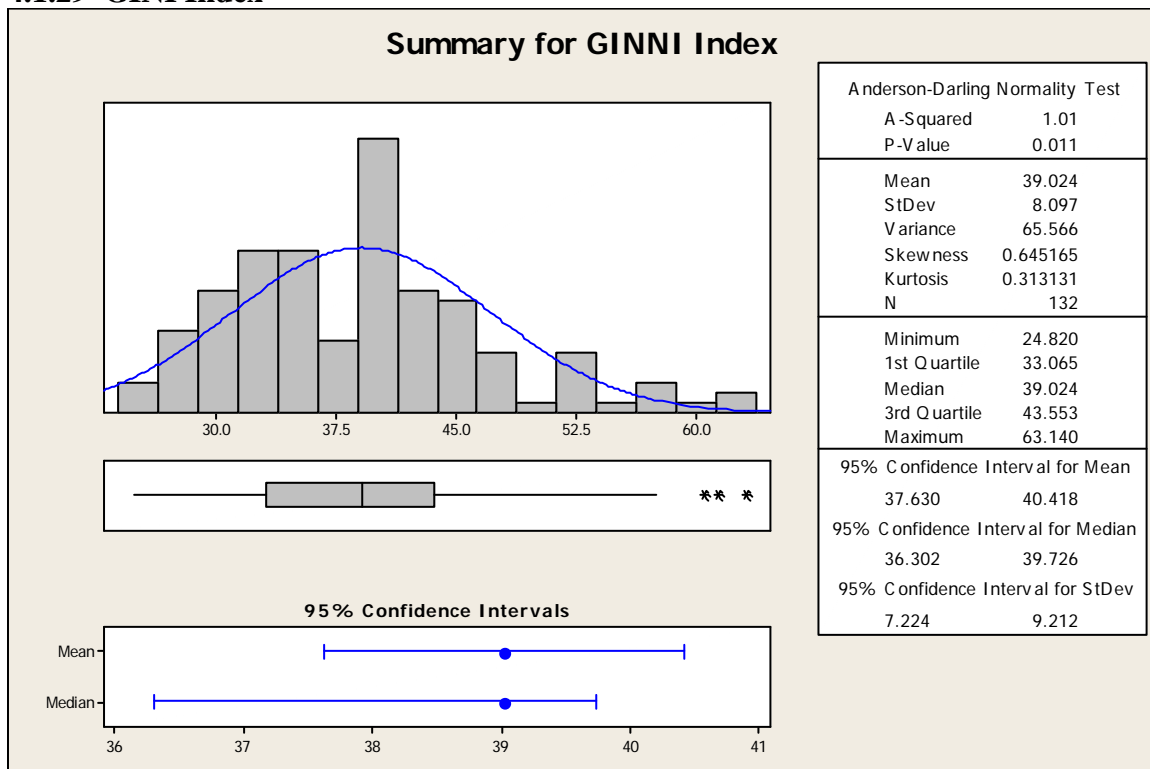


Figure 30: graph - GINI Index

Análise da variável:

GINI Index

Medida de desigualdade social utilizada para medir a desigualdade na distribuição de renda, ou desigualdade da riqueza.

A média tem valor de 39.024, a mediana de 39.024, e apresentam valores idênticos. A concentração dos dados no gráfico se encontra levemente a esquerda, o que denota uma situação aparentemente positiva.

Temos 3 *outliers*, são eles: Botswana (60.46), Namíbia (61.32) e África do Sul (63.14), e na outra extremidade os 3 países com resultados: Ucrânia (24.82), Suécia (26.08) e Eslovênia (24.87).

4.1.30 Democracy Index – 2014

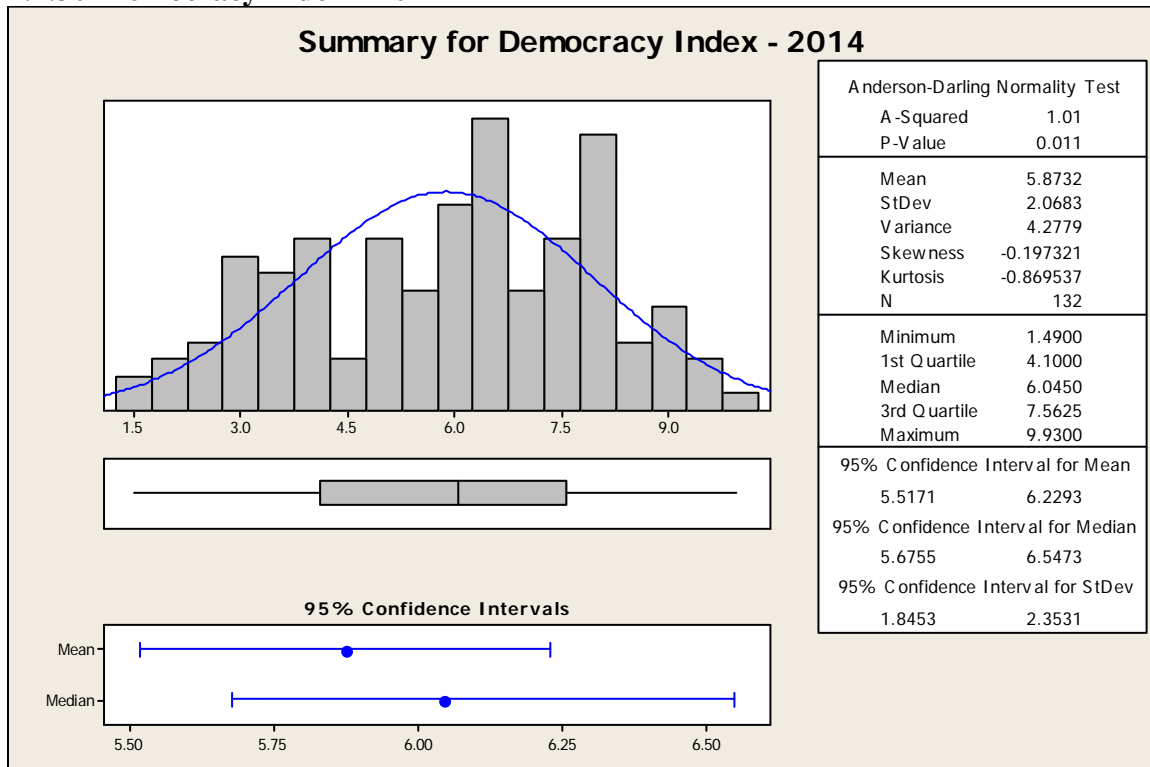


Figure 31: graph - Democracy Index – 2014

Análise da variável:

Democracy Index – 2014

A variável Índice de Democracia é formada por 5 categorias: Processo eleitoral e pluralismo, Liberdade civil, Funcionamento do governo, Participação política e Cultura política.

A curva da distribuição se aproxima de uma normal. A média tem valor de 5.8732, a mediana de 6.0450. No estrato entre 6.25 e 6.75, encontramos 17 países. A Noruega (9.93) é o país que apresenta o maior índice de Democracia.

Considerações sobre a estatística descritiva das principais variáveis

Os índices analisados a partir de dados de pesquisa de 132 países que refletem o grau de evolução e engajamento dos países no quesito humano, sustentável e natural. Os resultados indicam os países que de uma forma geral estão alinhados com seus discursos e com os relatórios de organismos como a UNICEF, o Banco Mundial, WEF – World Econômica Forum, FAO e outros órgãos preocupados em identificar a evolução dos povos do globo.

Indicadores como o IDH – Índice de Desenvolvimento Humano apresentado nesse trabalho, por exemplo, ao ser se analisa dados do IDH de países da África Sub-saariana, República centro-africana e Lesotho, vemos resultados das variáveis que fazem parte deste indicador como saúde, educação, uso de água potável, instalações sanitária que indicam a qualidade de vida, estão baixos quando comparados com outros países.

2. RELAÇÃO ENTRE AS VARIÁVEIS QUANTITATIVAS

2.1 Correlação linear

A matriz de correlação linear apresenta o teste de significância P-Value. Para a correlação foi utilizado o índice de Pearson. Conforme Las Casas e Guevara (2010)¹, a correlação (valores entre -1 e 1) é expressada pelo grau de relação linear que existe entre elas. Valores próximos a zero significam nenhuma relação, e próximos a 1 e -1 mostram alta relação positiva ou negativa, respectivamente. Vale ressaltar que o índice de correlação entre as variáveis não requer, necessariamente, que exista uma relação de causa-efeito entre as variáveis para dar sentido às relações. A seguir, apresentamos a matriz de correlação das variáveis estudadas no tema Biodiversidade - Capital Natural e Social:

Correlations: Social Progress, Basic Human Needs, Foundations Wellbeing, Opportu...

	Social Progress	Basic Human Needs	Foundations of W
Basic Human Needs	0.934 0.000		
Foundations of W	0.939 0.000	0.821 0.000	
Opportunity	0.922 0.000	0.748 0.000	0.847 0.000
Health and Welln	0.437 0.000	0.323 0.000	0.503 0.000
Personal Freedom	0.865 0.000	0.715 0.000	0.817 0.000
Greenhouse gas e	-0.342 0.000	-0.355 0.000	-0.319 0.000
Water withdrawal	0.140 0.109	0.288 0.001	0.001 0.994
Biodiversity and	0.353 0.000	0.184 0.034	0.566 0.000
Modern slavery,	-0.578 0.000	-0.564 0.000	-0.505 0.000
Satisfied demand	0.743 0.000	0.765 0.000	0.695 0.000
Women treated wi	0.162 0.064	0.201 0.021	0.109 0.215
Tolerance for im	0.243 0.005	0.066 0.454	0.205 0.018
Tolerance for ho	0.721	0.593	0.650

¹ LAS CASAS, Alexandre Luzzi; GUEVARA, Arnaldo José de Hoyos. Pesquisas de Marketing. São Paulo: Atlas, 2010.

	0.000	0.000	0.000
Discrimination a	-0.656 0.000	-0.496 0.000	-0.588 0.000
Religious tolera	0.196 0.024	0.053 0.547	0.198 0.023
Community safety	0.681 0.000	0.600 0.000	0.630 0.000
Years of tertiary	0.625 0.000	0.601 0.000	0.501 0.000
Women`s average	0.829 0.000	0.827 0.000	0.748 0.000
Inequality in th	-0.772 0.000	-0.755 0.000	-0.682 0.000
IDH - 2013	0.868 0.000	0.901 0.000	0.775 0.000
1. Governança	0.890 0.000	0.757 0.000	0.827 0.000
EPI Score	0.917 0.000	0.888 0.000	0.879 0.000
EV - Water Resources	0.765 0.000	0.725 0.000	0.675 0.000
EV - Agriculture	-0.215 0.013	-0.217 0.012	-0.173 0.048
EV - Forests	0.124 0.155	0.257 0.003	-0.015 0.867
EV - Fisheries	-0.027 0.758	-0.028 0.746	0.002 0.978
Happy Planet Ind	0.296 0.001	0.331 0.000	0.293 0.001
GINNI Index	-0.308 0.000	-0.415 0.000	-0.217 0.013
Democracy Index	0.789 0.000	0.604 0.000	0.768 0.000
	Opportunity	Health and Welln	Personal Freedom
Health and Welln	0.438 0.000		
Personal Freedom	0.907 0.000	0.504 0.000	
Greenhouse gas e	-0.273 0.002	-0.202 0.020	-0.235 0.007
Water withdrawal	0.036 0.685	0.039 0.658	0.079 0.368
Biodiversity and	0.330 0.000	0.308 0.000	0.348 0.000
Modern slavery,	-0.531 0.000	-0.264 0.002	-0.682 0.000

Satisfied demand	0.602 0.000	0.218 0.012	0.688 0.000
Women treated wi	0.121 0.168	0.175 0.045	0.155 0.076
Tolerance for im	0.435 0.000	0.333 0.000	0.391 0.000
Tolerance for ho	0.781 0.000	0.518 0.000	0.746 0.000
Discrimination a	-0.767 0.000	-0.264 0.002	-0.674 0.000
Religious tolera	0.328 0.000	0.117 0.183	0.285 0.001
Community safety	0.680 0.000	0.245 0.005	0.605 0.000
Years of tertiary	0.618 0.000	0.165 0.058	0.516 0.000
Women`s average	0.723 0.000	0.084 0.340	0.652 0.000
Inequality in th	-0.701 0.000	-0.047 0.593	-0.594 0.000
IDH - 2013	0.719 0.000	0.276 0.001	0.688 0.000
1. Governança	0.916 0.000	0.436 0.000	0.844 0.000
EPI Score	0.796 0.000	0.361 0.000	0.755 0.000
EV - Water Resou	0.724 0.000	0.332 0.000	0.655 0.000
EV - Agriculture	-0.200 0.021	0.044 0.615	-0.152 0.083
EV - Forests	0.042 0.634	-0.223 0.010	0.006 0.942
EV - Fisheries	-0.042 0.635	0.185 0.034	0.014 0.869
Happy Planet Ind	0.196 0.024	0.390 0.000	0.283 0.001
GINNI Index	-0.185 0.034	-0.008 0.924	-0.086 0.325
Democracy Index	0.869 0.000	0.450 0.000	0.758 0.000
Water withdrawal	Greenhouse gas e -0.173 0.048	Water withdrawal	Biodiversity and
Biodiversity and	0.049 0.574	-0.201 0.021	
Modern slavery,	0.171 0.050	-0.275 0.001	-0.141 0.107

Satisfied demand	-0.247 0.004	0.314 0.000	0.172 0.048
Women treated wi	-0.043 0.623	0.102 0.246	0.048 0.584
Tolerance for im	0.074 0.398	-0.203 0.020	0.173 0.047
Tolerance for ho	-0.195 0.025	0.067 0.444	0.272 0.002
Discrimination a	0.190 0.029	0.064 0.469	-0.299 0.000
Religious tolera	-0.052 0.551	-0.184 0.035	0.208 0.017
Community safety	-0.392 0.000	0.129 0.139	0.243 0.005
Years of tertiar	-0.186 0.033	0.211 0.015	0.104 0.233
Women`s average	-0.249 0.004	0.278 0.001	0.159 0.068
Inequality in th	0.201 0.021	-0.142 0.105	-0.173 0.048
IDH - 2013	-0.332 0.000	0.332 0.000	0.183 0.036
1. Governança	-0.283 0.001	0.012 0.892	0.355 0.000
EPI Score	-0.204 0.019	0.155 0.076	0.455 0.000
EV - Water Resou	-0.203 0.019	0.176 0.043	0.216 0.013
EV - Agriculture	0.131 0.133	-0.012 0.893	-0.063 0.471
EV - Forests	0.004 0.967	0.184 0.035	-0.309 0.000
EV - Fisheries	-0.053 0.544	0.025 0.777	0.040 0.647
Happy Planet Ind	-0.278 0.001	0.242 0.005	-0.024 0.781
GINNI Index	0.190 0.029	-0.037 0.672	0.120 0.172
Democracy Index	-0.307 0.000	-0.059 0.500	0.300 0.000
	Modern slavery,	Satisfied demand	Women treated wi
Satisfied demand	-0.580 0.000		
Women treated wi	0.037 0.672	-0.063 0.472	
Tolerance for im	-0.080	-0.099	0.091

	0.360	0.256	0.301
Tolerance for ho	-0.431 0.000	0.487 0.000	0.091 0.300
Discrimination a	0.325 0.000	-0.313 0.000	-0.137 0.117
Religious tolera	0.003 0.970	0.045 0.609	-0.129 0.140
Community safety	-0.409 0.000	0.501 0.000	0.054 0.542
Years of tertiary	-0.436 0.000	0.441 0.000	0.115 0.190
Women`s average	-0.554 0.000	0.773 0.000	0.067 0.448
Inequality in th	0.483 0.000	-0.634 0.000	-0.141 0.108
IDH - 2013	-0.580 0.000	0.765 0.000	0.111 0.207
1. Governança	-0.437 0.000	0.514 0.000	0.283 0.001
EPI Score	-0.553 0.000	0.683 0.000	0.205 0.018
EV - Water Resources	-0.456 0.000	0.498 0.000	0.326 0.000
EV - Agriculture	-0.028 0.751	-0.153 0.080	-0.175 0.045
EV - Forests	-0.028 0.754	0.146 0.095	0.125 0.152
EV - Fisheries	-0.017 0.849	0.035 0.689	-0.206 0.018
Happy Planet Ind	-0.291 0.001	0.502 0.000	-0.242 0.005
GINNI Index	0.141 0.106	-0.159 0.069	-0.323 0.000
Democracy Index	-0.372 0.000	0.511 0.000	-0.007 0.938

	Tolerance for im	Tolerance for ho	Discrimination a
Tolerance for ho	0.500 0.000		
Discrimination a	-0.428 0.000	-0.559 0.000	
Religious tolera	0.324 0.000	0.210 0.015	-0.506 0.000
Community safety	0.212 0.015	0.606 0.000	-0.454 0.000
Years of tertiary	0.151 0.084	0.482 0.000	-0.330 0.000

Women`s average	-0.018 0.837	0.469 0.000	-0.464 0.000
Inequality in th	-0.065 0.460	-0.443 0.000	0.483 0.000
IDH - 2013	0.035 0.693	0.590 0.000	-0.470 0.000
1. Governança	0.330 0.000	0.678 0.000	-0.737 0.000
EPI Score	0.202 0.020	0.644 0.000	-0.560 0.000
EV - Water Resou	0.201 0.021	0.601 0.000	-0.542 0.000
EV - Agriculture	0.153 0.080	-0.072 0.414	0.113 0.198
EV - Forests	0.047 0.591	-0.013 0.886	-0.006 0.943
EV - Fisheries	-0.018 0.839	0.040 0.650	0.057 0.516
Happy Planet Ind	-0.114 0.195	0.323 0.000	0.031 0.720
GINNI Index	0.120 0.169	-0.061 0.488	0.139 0.111
Democracy Index	0.264 0.002	0.637 0.000	-0.629 0.000
Community safety	Religious tolera 0.135 0.124	Community safety	Years of tertiary
Years of tertiary	0.002 0.984	0.442 0.000	
Women`s average	0.085 0.332	0.632 0.000	0.621 0.000
Inequality in th	-0.122 0.165	-0.595 0.000	-0.484 0.000
IDH - 2013	0.013 0.884	0.626 0.000	0.608 0.000
1. Governança	0.208 0.017	0.574 0.000	0.553 0.000
EPI Score	0.096 0.273	0.611 0.000	0.605 0.000
EV - Water Resou	0.020 0.823	0.527 0.000	0.603 0.000
EV - Agriculture	0.066 0.453	-0.215 0.013	-0.275 0.001
EV - Forests	-0.114 0.192	-0.023 0.791	0.098 0.263
EV - Fisheries	0.008 0.928	0.170 0.051	-0.077 0.378

Happy Planet Ind	-0.053 0.543	0.210 0.016	0.164 0.060
GINNI Index	0.206 0.018	-0.135 0.122	-0.309 0.000
Democracy Index	0.192 0.028	0.546 0.000	0.438 0.000

	Women`s average	Inequality in th	IDH - 2013
Inequality in th	-0.852 0.000		
IDH - 2013	0.844 0.000	-0.736 0.000	
1. Governança	0.650 0.000	-0.637 0.000	0.697 0.000
EPI Score	0.772 0.000	-0.722 0.000	0.824 0.000
EV - Water Resou	0.623 0.000	-0.560 0.000	0.690 0.000
EV - Agriculture	-0.188 0.030	0.186 0.033	-0.207 0.017
EV - Forests	0.210 0.016	-0.203 0.020	0.201 0.021
EV - Fisheries	-0.057 0.518	0.118 0.177	-0.024 0.785
Happy Planet Ind	0.274 0.001	-0.135 0.124	0.333 0.000
GINNI Index	-0.310 0.000	0.406 0.000	-0.297 0.001
Democracy Index	0.559 0.000	-0.542 0.000	0.577 0.000

	1. Governança	EPI Score	EV - Water Resou
EPI Score	0.811 0.000		
EV - Water Resources	0.783 0.000	0.840 0.000	
EV - Agriculture	-0.305 0.000	-0.188 0.031	-0.297 0.001
EV - Forests	0.096 0.272	0.200 0.022	0.127 0.147
EV - Fisheries	-0.073 0.404	-0.017 0.849	-0.111 0.206
Happy Planet Ind	0.089 0.312	0.197 0.024	0.086 0.325
GINNI Index	-0.247 0.004	-0.365 0.000	-0.421 0.000
Democracy Index	0.849 0.000	0.633 0.000	0.577 0.000

	EV - Agriculture	EV - Forests	EV - Fisheries
EV - Forests	-0.014 0.873		
EV - Fisheries	0.064 0.467	-0.117 0.181	
Happy Planet Ind	0.057 0.518	-0.035 0.692	0.192 0.028
GINNI Index	0.072 0.413	-0.175 0.045	0.137 0.116
Democracy Index	-0.249 0.004	-0.022 0.803	0.026 0.770
	Happy Planet Ind	GINNI Index	
GINNI Index	0.051 0.562		
Democracy Index	0.231 0.008	-0.118 0.176	

Cell Contents: Pearson correlation
P-Value

2.2 Gráficos de dispersão (*Scatterplot*)

Para efeito de análise prévia, foram construídos gráficos de dispersão evidenciando a relação entre as variáveis com maior grau de correlação. Como critério de escolha para elaboração dos gráficos, foi assumido os casos cujos resultados de correlação foram maior que 0.8, maior que -0.8 (negativo) e p-Value menor que 0.05.

Pelo tema central deste trabalho, buscamos destacar as correlações entre a variável *EPI Score* e outras de relevância. O índice *Environmental Performance Index (EPI)*, foi criado pela ONU – Organização das Nações Unidas, para atender o programa de Objetivos de Desenvolvimento do Milênio, em parceria com os centros de pesquisas das universidades de Yale e Columbia. O EPI pode ser considerado um método para quantificar a efetividade das políticas e ações voltadas às questões do Meio-Ambiente e seus impactos na saúde do ser humano, biodiversidade, sociedade e economia.

2.2.1 Variável de interesse *EPI Score* – Índice de Performance Ambiental

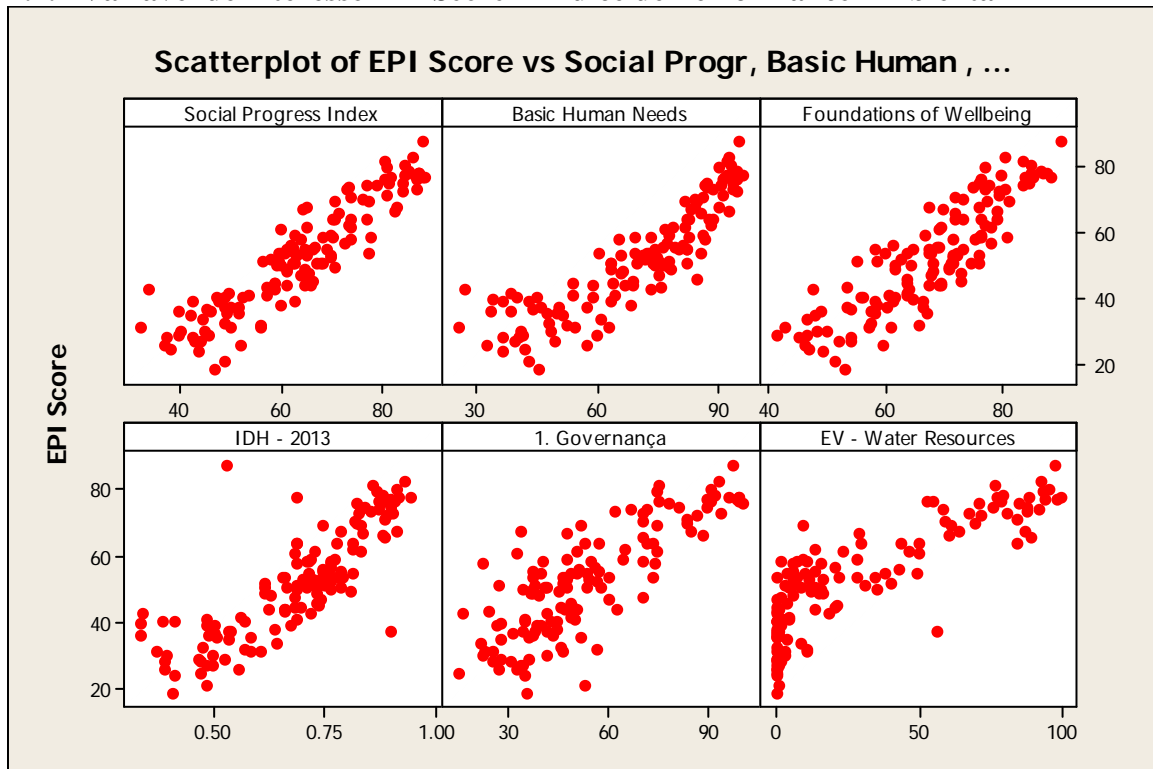


Figura 32: Gráficos de dispersão, variável de interesse *EPI Score* – Índice de Performance Ambiental

No gráfico acima (figura 32), percebe-se correlação linear positiva entre a variável EPI e as demais variáveis do estudo. O maior destaque foi com a variável Índice de Progresso Social (Social Progress Index), que apresentou correlação de 0.917 e p-Value de 0.000.

2.2.2 Variável de interesse *Social Progress Index* – Índice de Progresso Social

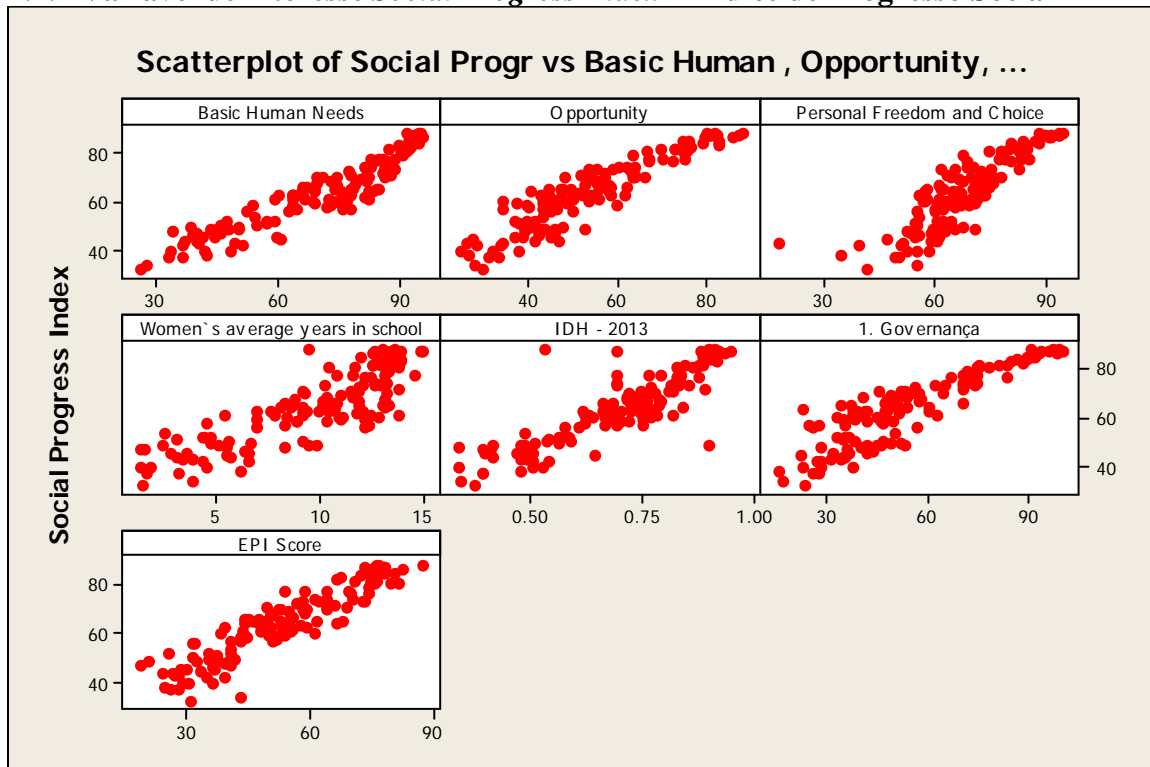


Figura 33: Gráficos de dispersão, variável de interesse *Social Progress Index* – Índice de Progresso Social

O gráfico baseado na variável de Índice de Progresso Social (*Social Progress Index*), mostra uma correlação positiva entre as variáveis. Com destaques para Necessidades Humanas Básicas (*Basic Human Needs*) com correlação de 0.934 e a variável Oportunidade (*Opportunity*) que apresentou correlação de 0.922 com a variável de referência.

5.2.3 Variável de interesse *Basic Human Needs* – *Necessidades Humanas Básicas*

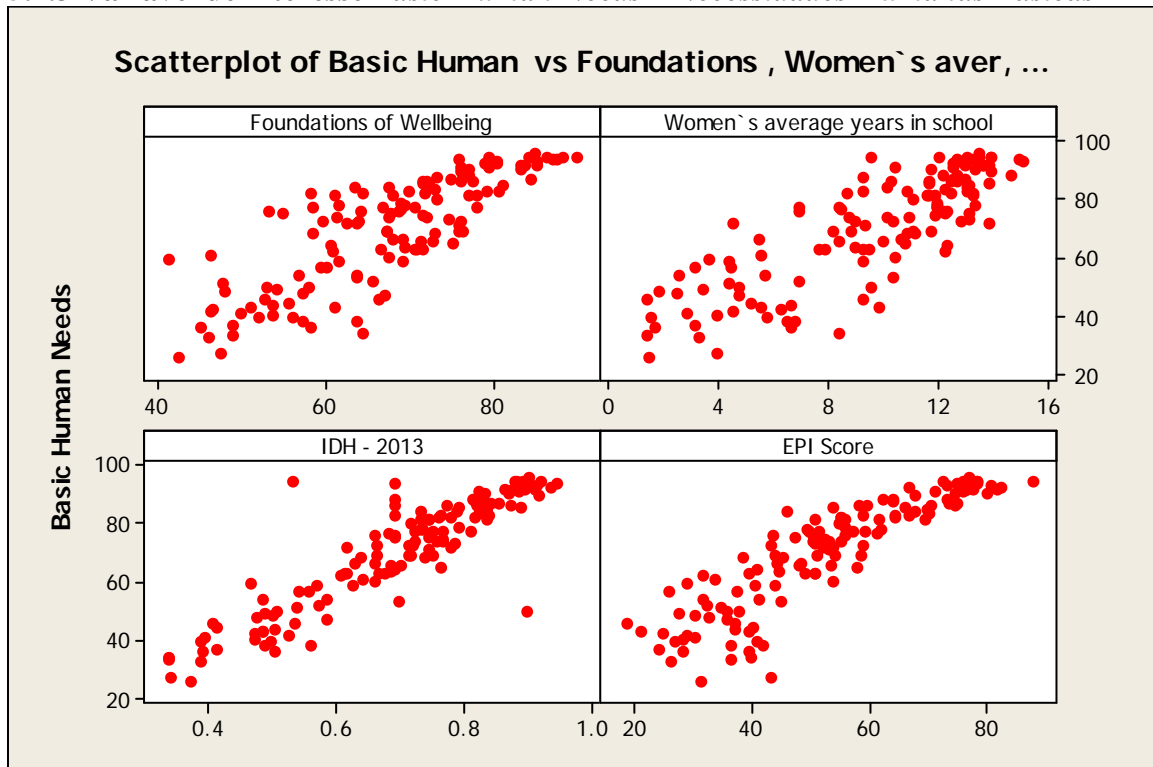


Figura 34: Gráficos de dispersão, variável de interesse *Basic Human Needs* – *Necessidades Humanas Básicas*

Acima nota-se correlação positiva entre a variável *Necessidades Humanas Básicas* (*Basic Human Needs*) e as demais verificadas. Destaca-se a variável IDH-2013, Índice de Desenvolvimento Humano, que resultou em 0.901 de grau de relação linear.

2.2.4 Variável de interesse *Foundations of Wellbeing* –*Fundamentos do Bem-Estar*

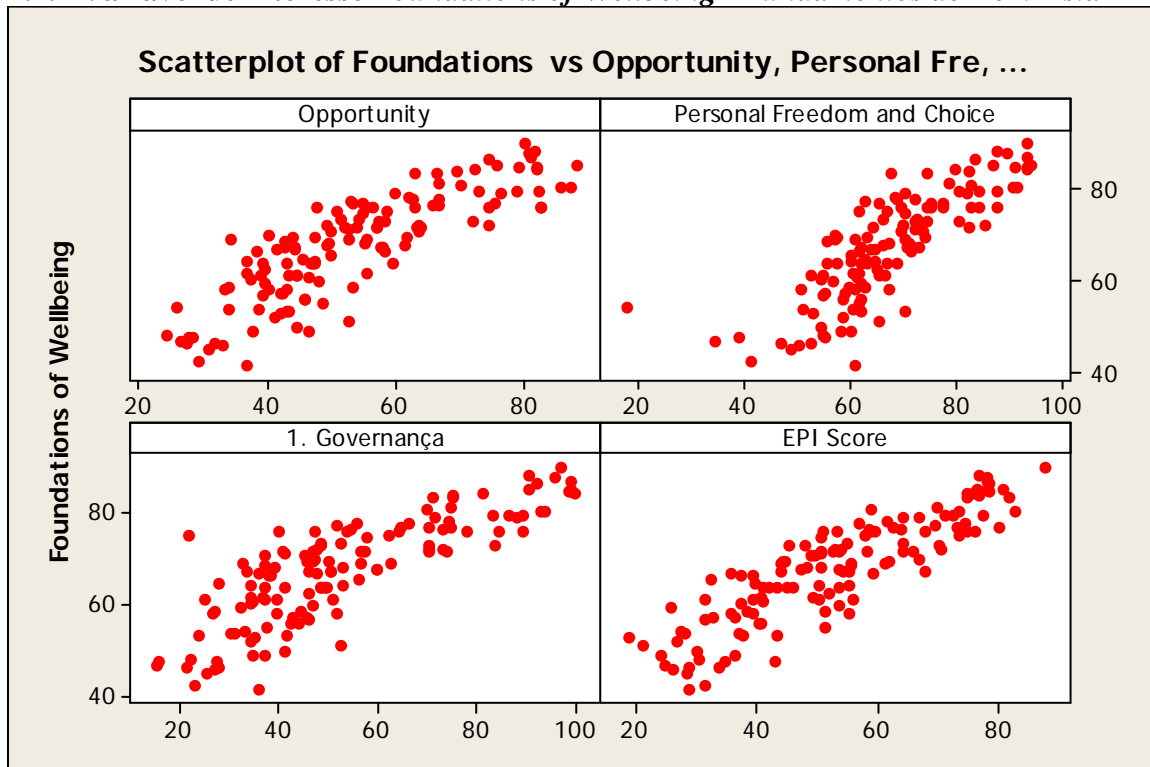


Figura 35: Gráficos de dispersão, variável de interesse *Foundations of Wellbeing* –*Fundamentos do Bem-Estar*

A variável Fundamentos do Bem-Estar (*Foundations of Wellbeing*), mostra no gráfico acima correlação positiva entre as variáveis em destaque no estudo. A que apresentou maior grau de correlação foi o Índice de Performance Ambiental (*EPI Score*) com grau de 0.879.

2.3 Linhas de tendência

Para esta análise, utilizamos a relação entre as variáveis com maior grau de correlação.

2.3.1 Variável de interesse *EPI Score* – Índice de Performance Ambiental

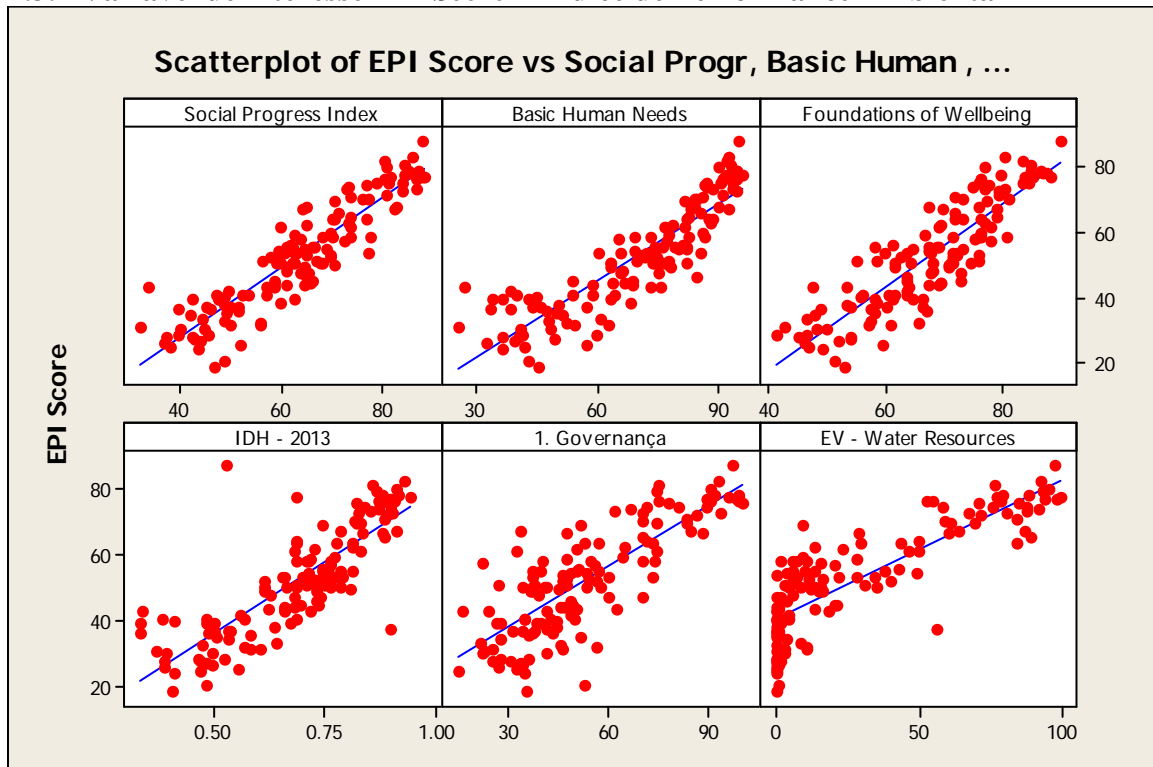


Figura 36: Gráficos de dispersão com linha de tendência, para *EPI Score* – Índice de Performance Ambiental

No gráfico acima (figura 36), percebe-se correlação linear positiva entre a variável EPI e as demais variáveis do estudo. O maior destaque foi com a variável Índice de Progresso Social (*Social Progress Index*), que apresentou correlação de 0.917 e *p-Value* de 0.000.

2.3.2 Variável de interesse *Social Progress Index* – Índice de Progresso Social

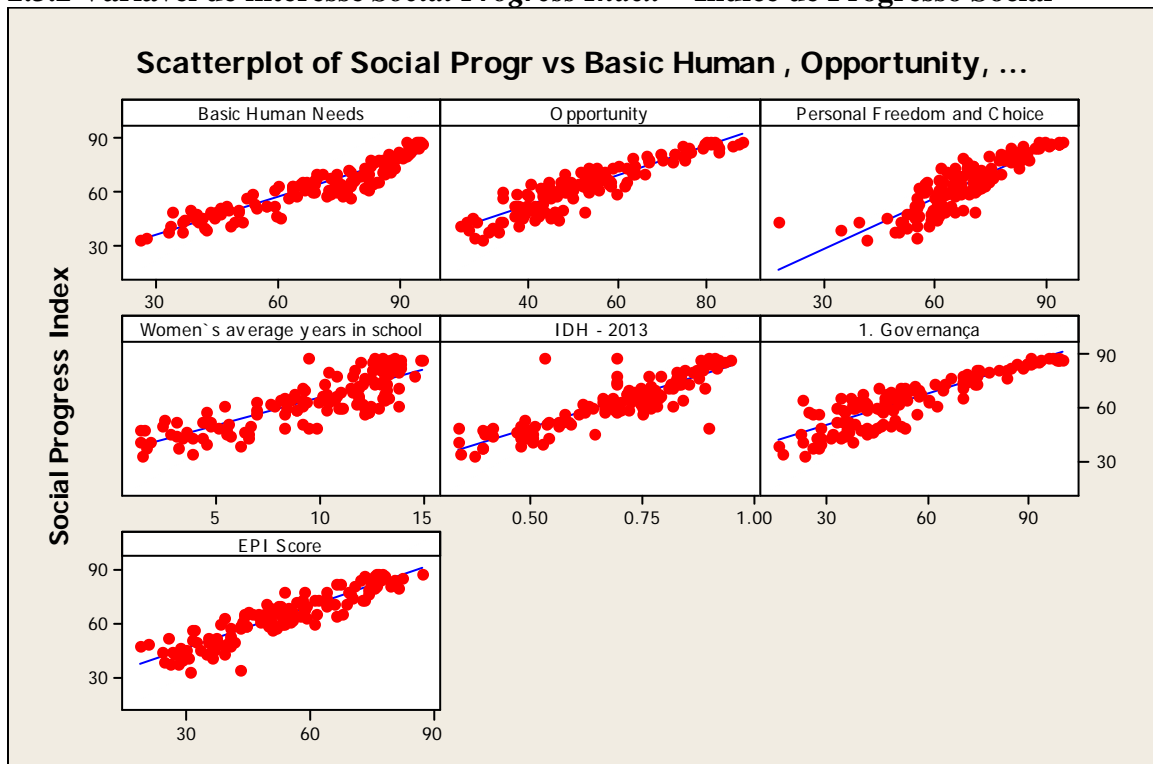


Figura 37: Gráficos de dispersão com linha de tendência para *Social Progress Index* – Índice de Progresso Social

O gráfico baseado na variável de Índice de Progresso Social (*Social Progress Index*), mostra uma correlação positiva entre as variáveis. Com destaques para as variáveis Necessidades Humanas Básicas (*Basic Human Needs*) com correlação de 0.934 e Oportunidade (*Opportunity*) que apresentou correlação de 0.922 com a variável de referência.

2.3.3 Variável de interesse *Basic Human Needs* – Necessidades Básicas Humanas

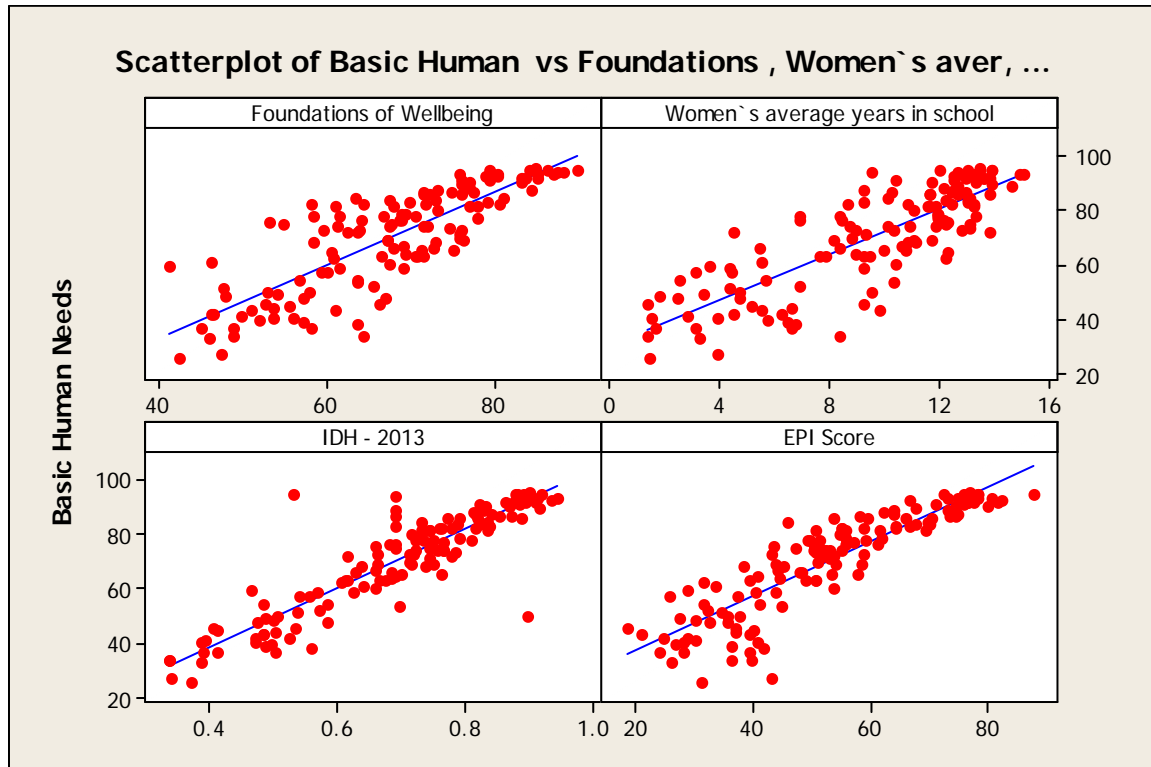


Figura 38: Gráficos de dispersão com linha de tendência para *Basic Human Needs* – Necessidades Básicas Humanas

Acima nota-se correlação positiva entre a variável Necessidades Humanas Básicas (*Basic Human Needs*) e as demais verificadas. Destaca-se a variável IDH-2013, Índice de Desenvolvimento Humano, que resultou em 0.901 de grau de relação linear.

2.3.4 Variável de interesse *Foundations of Wellbeing* – Fundamentos para o Bem-Estar

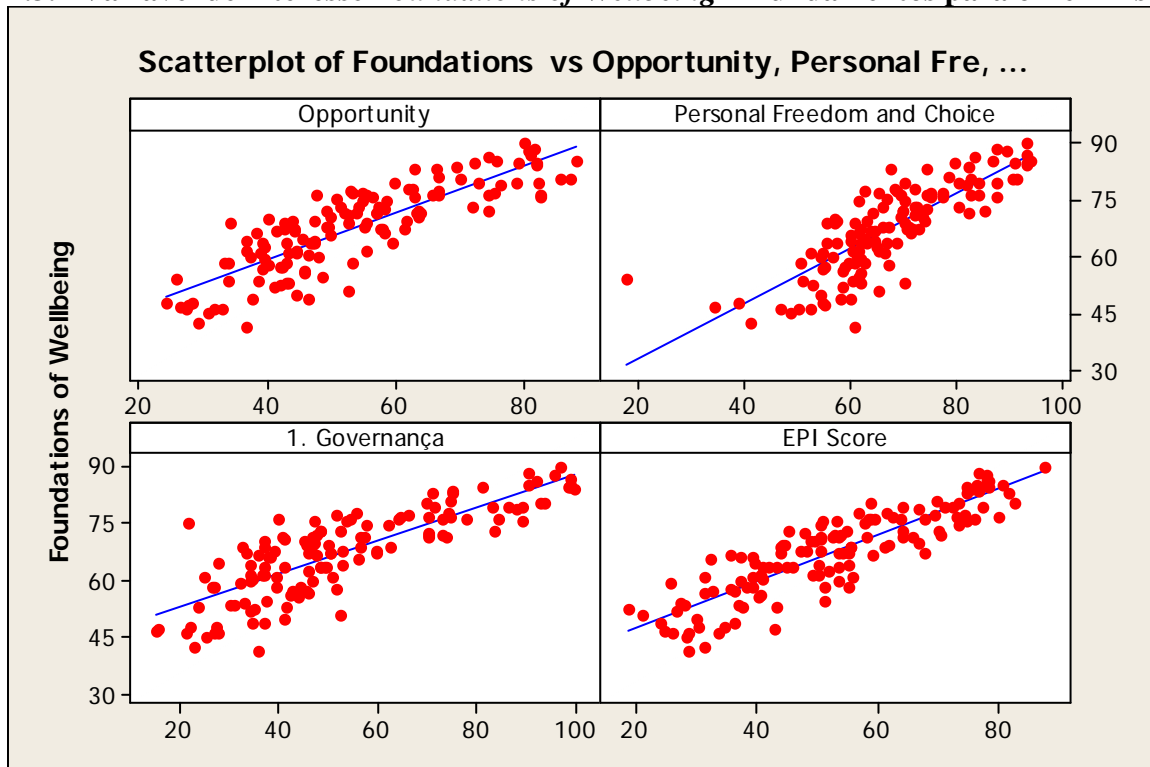


Figura 39: Gráficos de dispersão com linha de tendência para *Foundations of Wellbeing* – Fundamentos para o Bem-Estar

A variável Fundamentos do Bem-Estar (*Foundations of Wellbeing*), mostra no gráfico acima correlação positiva entre as variáveis em destaque no estudo. A que apresentou maior grau de correlação foi o Índice de Performance Ambiental (*EPI Score*) com grau de 0.879.

2.4 Dendrogram (Dendrograma ou Dendograma) – Diagrama de Árvore

Um Dendrograma (dendr(o) = árvore) é um tipo específico de diagrama ou representação icônica que organiza determinados fatores e variáveis. É um diagrama de similaridade. A interpretação de um dendrograma de similaridade entre amostras fundamenta-se na intuição: duas amostras próximas devem ter também valores semelhantes para as variáveis medidas. Ou seja, elas devem ser próximas matematicamente no espaço multidimensional. Portanto, quanto maior a proximidade entre as medidas relativas às amostras, maior a similaridade entre elas. O dendrograma hierarquiza esta similaridade de modo que podemos ter uma visão bidimensional da similaridade ou dissimilaridade de todo o conjunto de amostras utilizado no estudo. Segue abaixo o Dendrograma das variáveis analisadas:

Um dendrograma ilustra as informações na tabela de amalgamação na forma de um diagrama de árvore.

Por padrão, o nível de similaridade é medido no eixo vertical (também é possível exibir o nível de distância) e as diferentes observações são listadas ao longo do eixo horizontal. O gráfico mostra como os agrupamentos são formados: unindo duas observações individuais ou pareando uma observação individual com um agrupamento existente. É possível ver em que nível de similaridade os agrupamentos são formados, e a composição dos agrupamentos da partição final.

Para alguns conjuntos de dados, os métodos média, centróide, mediana e Ward podem não gerar um dendrograma hierárquico. Ou seja, as distâncias de amalgamação nem sempre aumentam com cada etapa. No dendrograma, essa etapa geraria uma junção para baixo no lugar de para cima.

O dendrograma indica as variáveis mais próximas em relação ao coeficiente de correlação ou alto grau de similaridade.

2.4.1 Análise de cluster das variáveis - *Dendrogram*

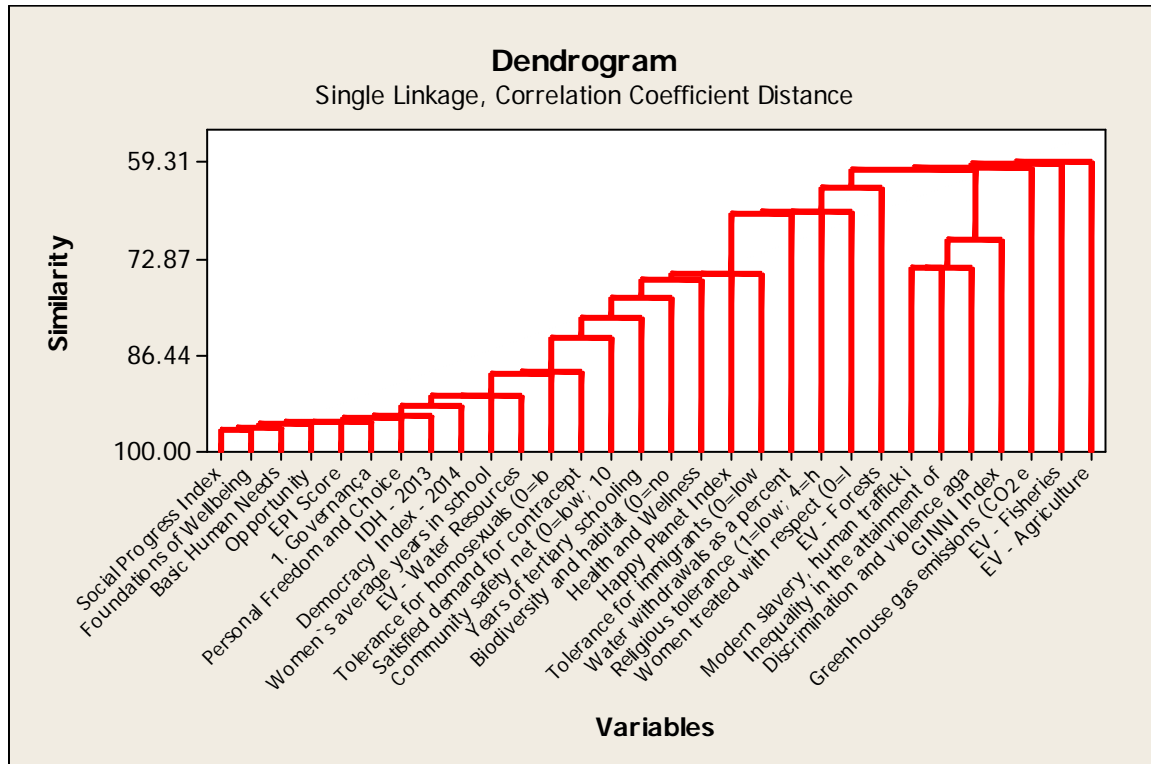


Gráfico (acima) gerado considerando apenas 1 cluster.

Em seguida, segue análise de cluster das variáveis deste estudo:

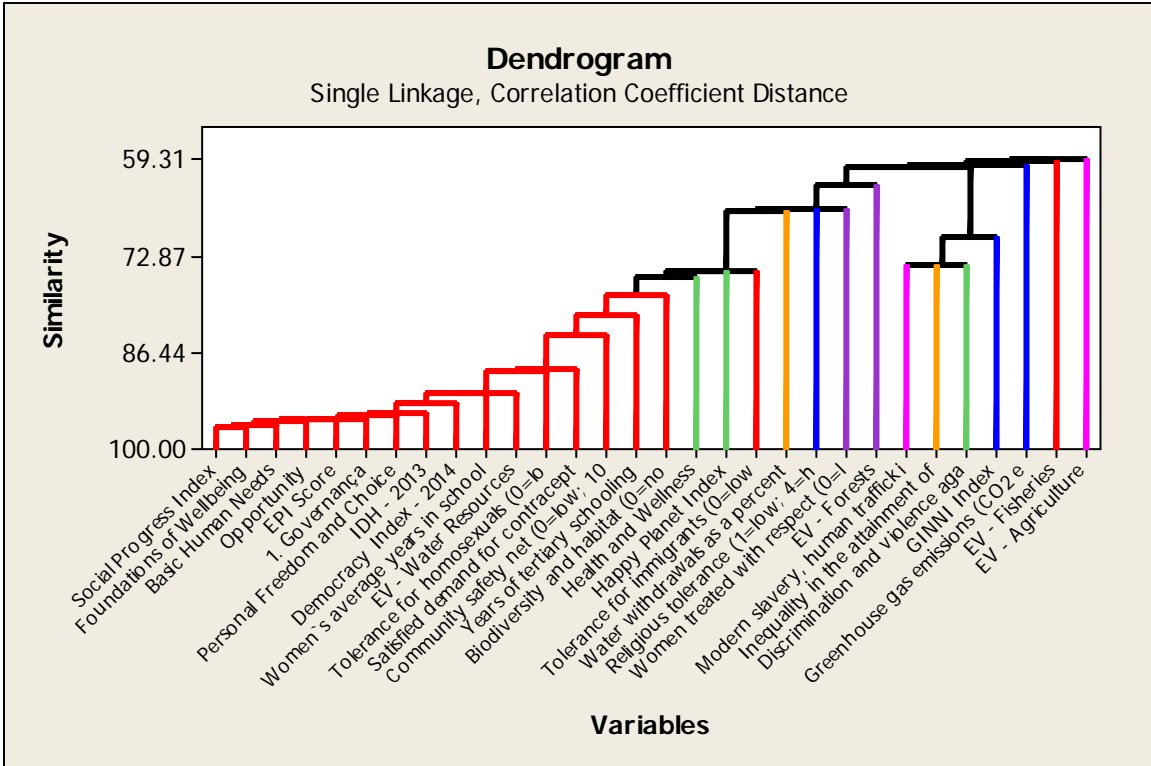
Cluster Analysis of Variables: Social Progress, Basic Human Needs, ...

Correlation Coefficient Distance, Single Linkage
Amalgamation Steps

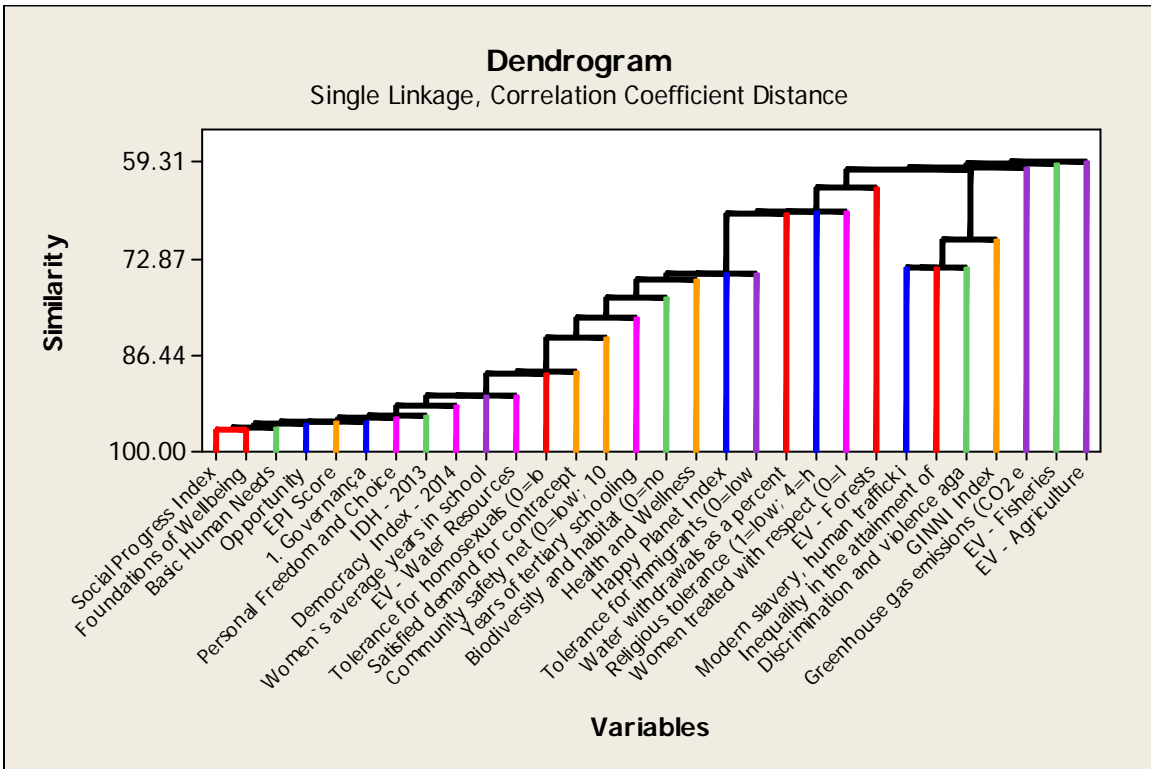
Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	29	96.9544	0.060912	1 3	1	2
2	28	96.6963	0.066073	1 2	1	3
3	27	96.0972	0.078056	1 4	1	4
4	26	95.8701	0.082598	1 23	1	5
5	25	95.8063	0.083875	1 22	1	6
6	24	95.3580	0.092840	1 6	1	7
7	23	95.0464	0.099071	1 21	1	8
8	22	93.4716	0.130567	1 30	1	9
9	21	92.1988	0.156024	1 19	1	10
10	20	92.0028	0.159944	1 24	1	11
11	19	89.0470	0.219061	1 14	1	12
12	18	88.6723	0.226554	1 11	1	13
13	17	84.0626	0.318747	1 17	1	14
14	16	81.2353	0.375293	1 18	1	15
15	15	78.2833	0.434334	1 9	1	16
16	14	75.8883	0.482233	1 5	1	17
17	13	75.0977	0.498045	1 28	1	18
18	12	74.9784	0.500432	1 13	1	19
19	11	74.1538	0.516925	10 20	10	2
20	10	74.1254	0.517492	10 15	10	3
21	9	70.3083	0.593835	10 29	10	4
22	8	66.5831	0.668338	1 8	1	20
23	7	66.3931	0.672139	1 16	1	21
24	6	66.3052	0.673897	1 12	1	22
25	5	62.8294	0.743413	1 26	1	23
26	4	60.3164	0.793671	1 10	1	27
27	3	60.0636	0.798727	1 7	1	28
28	2	59.5788	0.808425	1 27	1	29
29	1	59.3122	0.813756	1 25	1	30

Dendrogram

O gráfico abaixo utiliza um total de 15 clusters, divide as observações no ponto médio nível de similaridade (78.2833):



O seguinte gráfico foi construído considerando um total de 29 clusters, que segundo o software estatístico utilizado neste trabalho, é considerado como o maior nível de similaridade entre as variáveis (96.9544):



3. REGRESSÃO LINEAR MÚLTIPLA

3.1 Normalização dos Dados

Antes de partir para as análises de regressão, vamos proceder com a normalização de nossa base de dados.

A normalização dos dados permite atribuir uma nova escala para que todas as variáveis sejam expressas no mesmo padrão.

Neste trabalho foi utilizada a Normalização Min-Max, que se baseia nos valores mínimos e máximos das observações. Este método preserva o relacionamento entre os valores dos dados originais.

Todas as variáveis quantitativas neste trabalho foram normalizadas. A equação utilizada para normalização da base de dados foi:

$$(c4 - \text{MIN}(c4)) / (\text{MAX}(c4) - \text{MIN}(c4)) * 100$$

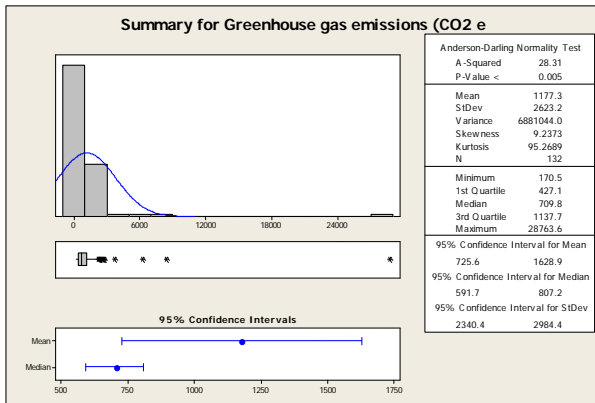
Onde: c4 refere-se a coluna dos dados que serão normalizados pelo Método Min-Mix

O procedimento acima, foi aplicado nas 30 variáveis quantitativas deste estudo. Adicionalmente, algumas variáveis tiveram ajustes em suas escalas (positivadas). Nas variáveis que demonstraram algum tipo de agravamento (negativo) a medida de seu aumento, por exemplo, o Índice de Emissão de Gases Poluentes que quanto maior seu resultado, menor é seu benefício ao meio ambiente e a saúde humana, foi aplicado a fórmula: $(100 - \text{"valor da célula"})$, para todas observações daquela variável. Desta maneira, todos os indicadores referentes a este estudo passam a refletir um melhor desempenho a medida que se aproximam de 100 e o inverso também é verdadeiro.

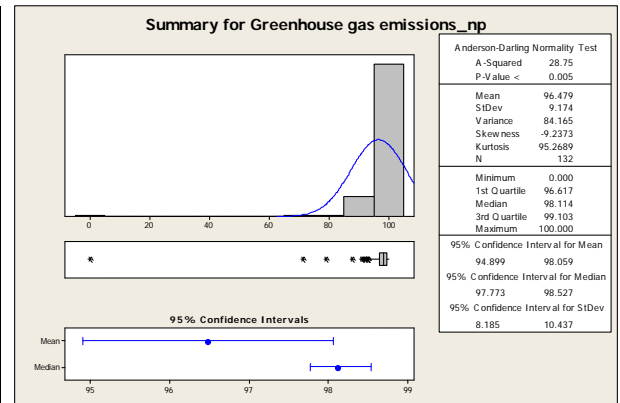
As seguintes variáveis foram "positivadas" para que permitissem o entendimento de quanto maior o seu valor, melhor (ou mais positiva) era a situação do indicador: *Greenhouse gas emissions_np*, *Water withdrawals_np*, *Modern slavery_np*, *Discrimination and violence_np*, *Inequality in the attainment_np*, *EV-Fisheries_np* e *GINI Index_np*.

Visualmente, estas variáveis são identificadas pelo sufixo "np", que denota ser uma variável normalizada e positivada.

Para efeitos de validação do método de "positivar" as variáveis deste estudo e garantir que não haveria perda de referência, foi gerado novo gráfico para a variável "*Greenhouse gas emissions_np*" e confrontado com o gráfico gerado na primeira fase deste trabalho.



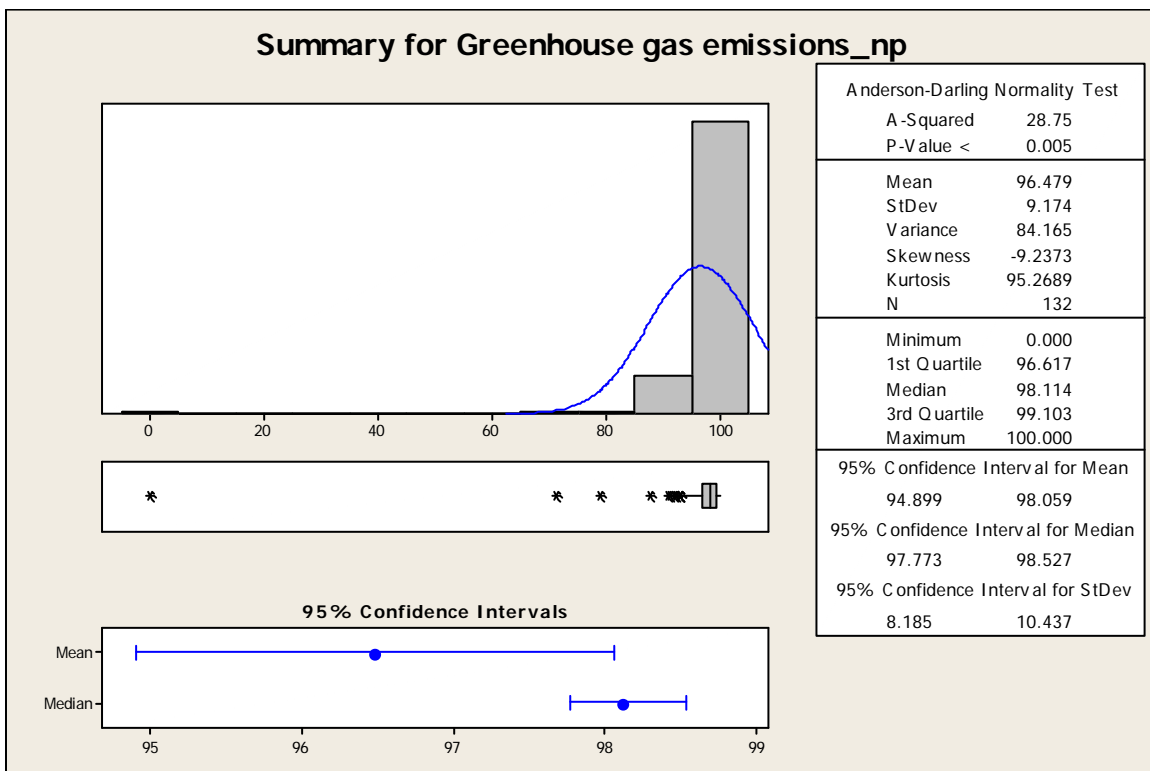
ANTES de normalizar e positivar

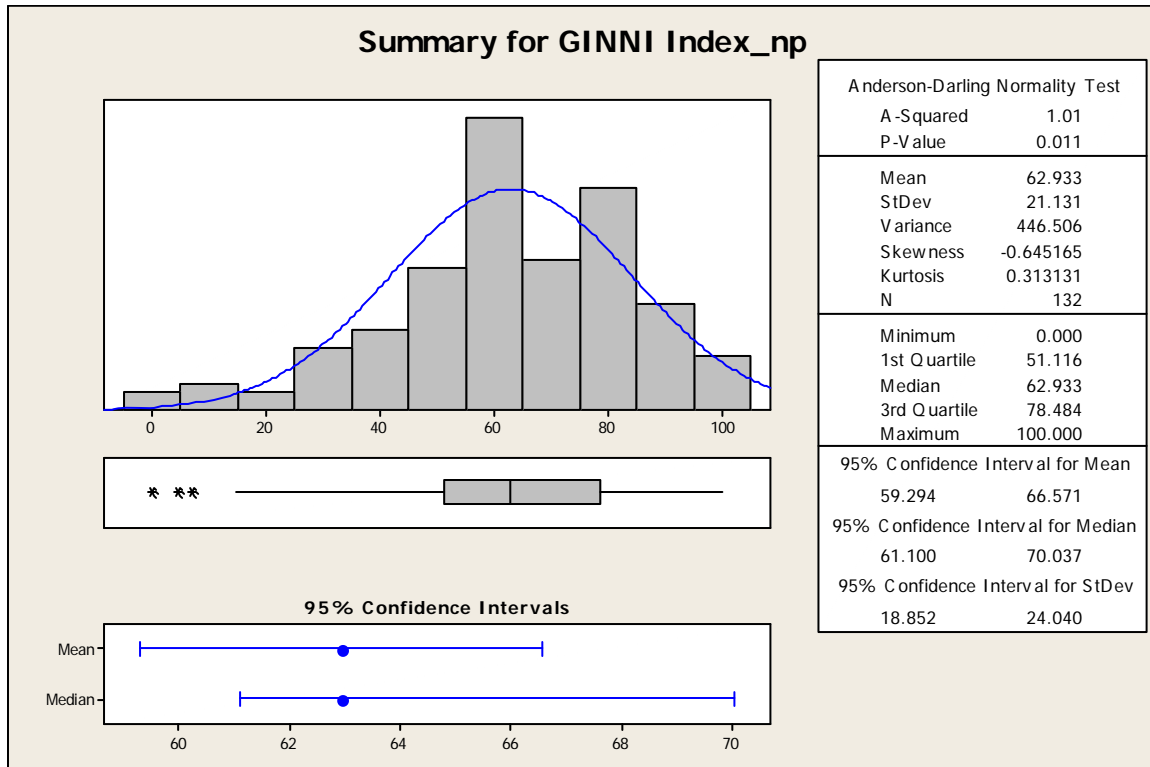


DEPOIS de normalizar e positivar

A comparação acima dos dois momentos, antes e depois de normalizar e positivar, mostra que foi mantido proporcionalmente o relacionamento na disposição dos dados.

Greenhouse gas emissions - variável NORMALIZADA e POSITIVADA – resultado





3.2 Análise de Regressão

Nesta etapa do trabalho, serão executadas as análises de regressão, para tanto será utilizado o software estatístico Minitab e a base dados contendo as variáveis referentes ao tema Biodiversidade – Capital Natural e Social, objeto deste estudo, já normalizadas e positivadas.

No software estatístico, a opção para executar a função de análise de regressão no menu é: *stat -> Regression -> Regression*

Será utilizada a variável *EPI Score* como variável independente (*response*), e como dependentes (preditoras), as variáveis: *Social Progress Index*, *Basic Human Needs*, *Foundations of Wellbeing*, *Opportunity*, 1. Governança, *Personal Freedom and Choice*, IDH – 2013, *Democracy Index – 2014*, *Women`s average years in school* e *EV - Water Resources.*, que apresentaram bom nível de similaridade, conforme dendrograma apresentado na seção anterior.

Abaixo, os resultados da regressão com as variáveis mencionadas:

Regression Analysis: EPI Score_n versus Social Progress, Basic Human Needs, ...

The regression equation is

$$\begin{aligned} \text{EPI Score}_n = & 344 + 90 \text{ Social Progress Index}_n - 37.5 \text{ Basic Human Needs}_n \\ & - 25.8 \text{ Foundations of Wellbeing}_n - 34.4 \text{ Opportunity}_n \\ & + 0.0200 \text{ Governança}_n - 0.0497 \text{ Personal Freedom and Choice}_n \\ & - 0.0171 \text{ IDH-2013}_n - 0.177 \text{ Democracy Index-2014}_n \\ & + 0.0036 \text{ Women`s years in school}_n + 0.257 \text{ EV-Water Resources}_n \end{aligned}$$

Predictor	Coef	SE Coef	T	P	
Constant	343.6	409.5	0.84	0.403	
Social Progress Index_n	90.4	109.9	0.82	0.412	<- descartar
Basic Human Needs_n	-37.54	45.96	-0.82	0.416	<- descartar
Foundations of Wellbeing_n	-25.77	31.90	-0.81	0.421	<- descartar
Opportunity_n	-34.44	41.94	-0.82	0.413	<- descartar
Governança_n	0.02005	0.07622	0.26	0.793	<- descartar
Personal Freedom and Choice_n	-0.04970	0.09268	-0.54	0.593	<- descartar
IDH-2013_n	-0.01712	0.06032	-0.28	0.777	<- descartar
Democracy Index-2014_n	-0.17736	0.05849	-3.03	0.003	
Women`s years in school_n	0.00359	0.04809	0.07	0.941	<- descartar
EV-Water Resources_n	0.25673	0.03308	7.76	0.000	

S = 6.95727 **R-Sq = 92.2%** R-Sq(adj) = 91.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	10	68997.8	6899.8	142.55	0.000
Residual Error	121	5856.8	48.4		
Total	131	74854.7			

Source	DF	Seq SS
Social Progress Index_n	1	62999.6
Basic Human Needs_n	1	554.4
Foundations of Wellbeing_n	1	782.4
Opportunity_n	1	3.4
Governança_n	1	363.1
Personal Freedom and Choice_n	1	9.0
IDH-2013_n	1	41.7
Democracy Index-2014_n	1	1327.9
Women`s years in school_n	1	0.2
EV-Water Resources_n	1	2916.1

Unusual Observations

Obs	Social Progress Index_n	EPI Score_n	Fit	SE Fit	Residual	St Resid
9	35	10.370	27.320	1.736	-16.950	-2.52R
23	3	35.399	12.282	2.268	23.117	3.51R
44	42	19.700	35.687	1.507	-15.987	-2.35R
77	26	0.000	16.864	2.112	-16.864	-2.54R
78	19	12.652	23.061	3.942	-10.409	-1.82 X
91	13	25.780	11.364	2.000	14.416	2.16R
106	68	73.224	56.749	1.627	16.475	2.44R
113	29	27.325	38.158	3.953	-10.832	-1.89 X
115	100	100.000	95.588	3.971	4.412	0.77 X
116	42	18.645	39.131	2.355	-20.486	-3.13R

R denotes an observation with a large standardized residual.

X denotes an observation whose X value gives it large leverage.

No resultado acima vemos que apesar do grau de similaridade no gráfico de Dendrograma e r-quadrado de 92.2% indicar uma boa relação entre as variáveis testadas, o resultado da análise mostrou o contrário, das 10 variáveis testadas, apenas 2 apresentaram p-value abaixo de 0.05 e destas, somente 1 com estatística t (de Student) com valor positivo.

Foi executado novamente a análise de regressão, desta vez mantido *EPI Score* como variável independente, descartado aquelas que apresentaram alto grau de imprecisão (p-value > 0.05) e considerado como dependentes (preditoras), as variáveis já testadas: *Democracy Index-2014* e *EV-Water Resources* e adicionado a variável: *Biodiversity and Habitat*.

Abaixo, os resultados com a nova constituição de variáveis testadas:

Regression Analysis: EPI Score_n versus Democracy Index, EV-Water Resources,...

The regression equation is

$$\text{EPI Score}_n = 14.0 + 0.151 \text{ Democracy Index-2014}_n + 0.502 \text{ EV-Water Resources}_n + 0.222 \text{ Biodiversity and habitat}_n$$

Predictor	Coef	SE Coef	T	P
Constant	14.039	2.716	5.17	0.000
Democracy Index-2014_n	0.15105	0.04842	3.12	0.002
EV-Water Resources_n	0.50182	0.03503	14.33	0.000
Biodiversity and habitat_n	0.22174	0.03559	6.23	0.000

S = 10.8256 **R-Sq = 80.0%** R-Sq(adj) = 79.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	59854	19951	170.24	0.000
Residual Error	128	15001	117		
Total	131	74855			

Source	DF	Seq SS
Democracy Index-2014_n	1	30003
EV-Water Resources_n	1	25303
Biodiversity and habitat_n	1	4548

Unusual Observations

Obs	Democracy Index-2014_n	EPI Score_n	Fit	SE Fit	Residual	St Resid
54	6	47.155	24.540	2.065	22.615	2.13R
77	51	0.000	24.905	2.047	-24.905	-2.34R
106	62	73.224	49.970	1.945	23.254	2.18R
113	19	27.325	49.008	2.732	-21.683	-2.07R
125	14	78.683	71.542	3.422	7.141	0.70 X

R denotes an observation with a large standardized residual.

X denotes an observation whose X value gives it large leverage.

Na segunda execução da análise de regressão, tomando as variáveis pré-selecionadas, temos bom nível de similaridade nos resultados, com *p-value* próximo de 0 e *r-quadrado* de 80%.. Todos os resultados da estatística *t* (de Student) apresentaram valor positivo.

3.3 Correção dos Dados (outliers)

A ferramenta de análise de regressão identificou um total de 5 observações em nossa base, as quais apresentaram valores inesperados (*outliers*) para a variável *EPI Score*, as referidas células foram modificadas manualmente, recebendo cada uma seu valor correspondente da coluna *FIT*, conforme abaixo:

Unusual Observations

Obs	Democracy Index-2014_n	EPI Score_n	Fit	SE Fit	Residual	St Resid
54	6	47.155	24.540	2.065	22.615	2.13R
77	51	0.000	24.905	2.047	-24.905	-2.34R
106	62	73.224	49.970	1.945	23.254	2.18R
113	19	27.325	49.008	2.732	-21.683	-2.07R
125	14	78.683	71.542	3.422	7.141	0.70 X

Aplicado os ajustes, executado novamente a análise, surgiram outras 5 observações com necessidade de correção. Inclusive repetindo a Obs nr.125

Unusual Observations

Obs	Democracy Index-2014_n	EPI Score_n	Fit	SE Fit	Residual	St Resid
9	51	10.370	30.912	1.368	-20.542	-2.08R
52	76	18.486	40.302	1.905	-21.816	-2.23R
70	61	3.437	25.429	2.437	-21.992	-2.28R
104	4	69.656	48.286	2.775	21.370	2.23R
125	14	71.542	70.164	3.150	1.378	0.15 X

3.4 Regressão Stepwise

A regressão stepwise é uma ferramenta automática usada nos estágios exploratórios da construção de modelos para identificar um subconjunto útil de preditores. O processo sistematicamente adiciona a variável mais significativa ou remove a variável menos significativa durante cada etapa.

Fonte:<http://support.minitab.com/pt-br/minitab/17/topic-library/modeling-statistics/regression-and-correlation/basics/basics-of-stepwise-regression/>

Devido ao problema de ajuste na fase anterior (seção 3.3), então somente para demonstrar o procedimento e fins de ilustração, será executado a ferramenta de regressão stepwise, utilizando a variável de interesse: *EPI Score*, e as seguintes variáveis como

dependentes: faixa de 'Social Progress Index_n'- até 'Governança_n' e a faixa de variáveis de 'EV-Water Resources_n'- até 'Democracy Index-2014_n'. Os resultados, seguem abaixo:

Stepwise Regression: EPI Score_n versus Social Progress, Basic Human Needs, ...

Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15

Response is EPI Score_n on 29 predictors, with N = 132

Step	1	2	3	4	5	6
Constant	2.1011	9.6294	5.2423	0.6514	-2.1860	-1.4417
Social Progress Index_n	0.842	0.569	0.506	0.173	0.221	0.408
T-Value	25.96	14.36	13.55	2.38	3.15	4.09
P-Value	0.000	0.000	0.000	0.019	0.002	0.000
EV-Water Resources_n		0.276	0.289	0.288	0.287	0.314
T-Value		9.03	10.48	11.44	12.02	12.26
P-Value		0.000	0.000	0.000	0.000	0.000
Biodiversity and habitat_n			0.126	0.177	0.199	0.195
T-Value			5.57	7.74	8.91	8.93
P-Value			0.000	0.000	0.000	0.000
Basic Human Needs_n				0.320	0.253	0.160
T-Value				5.17	4.16	2.30
P-Value				0.000	0.000	0.023
EV-Forests_n					0.079	0.087
T-Value					3.95	4.41
P-Value					0.000	0.000
Governança_n						-0.139
T-Value						-2.58
P-Value						0.011
S	9.48	7.45	6.71	6.12	5.80	5.67
R-Sq	83.83	90.09	92.02	93.41	94.14	94.43
R-Sq(adj)	83.71	89.94	91.84	93.20	93.91	94.17
Step	7	8	9			
Constant	11.68	10.45	14.25			

4. ANÁLISE DE TENDÊNCIAS

Nesta sessão será efetuada análise de tendência e projeções das variáveis temporais e quantitativas “*Terrestrial and marine protected areas*” (áreas terrestres e marinhas sob proteção), “*Forest area*” (área florestal) e “*Population*” (população total de um país), relacionadas ao tema **Biodiversidade - Capital Natural e Social**. Os dados se referem aos países Brasil, México, Portugal e Espanha e foram extraídos do site *Knoema*, que trata-se de um projeto de integração de bases de dados de diversos órgãos internacionais, como Banco Mundial, ONU, Unesco, entre outros.

A análise de cada variável está dividida em três partes. A primeira refere-se a análise do comportamento histórico através de gráficos. A segunda trata da análise de tendências, buscando encontrar a função estatística que melhor se adapte à cada uma das séries de variáveis analisadas. E a terceira busca a elaboração e análise de projeções através de extrapolações estatísticas. O software estatístico utilizado é o MINITAB.

4.1 Uma breve descrição do Projeto KNOEMA

O Projeto Knoema pode ser descrito como uma plataforma de conhecimento. A ideia de concepção reside no tripé: Disponibilização dos dados, Ferramentas de análise *online* e Ferramentas para elaboração de relatórios e apresentação dos resultados. Isso possibilita ao pesquisador selecionar um conjunto de dados de interesse, executar análises em busca de relevância e elaborar o relatório final com suas conclusões, diretamente na plataforma Knoema. fonte: <http://knoema.com/about/us>

4.2. Séries históricas

Os dados são séries históricas referentes ao Brasil, México, Portugal e Espanha, portanto, tratam-se de séries temporais. As séries vão de 1990 a 2012 para a variável “*Terrestrial and marine protected areas*” (áreas terrestres e marinhas sob proteção), e de 1990 a 2013 para as variáveis “*Forest area*” (área florestal) e “*Population*” (população total de um país).

Cabe mencionar que para a variável “*Terrestrial and marine protected areas*” (áreas terrestres e marinhas sob proteção), não havia dados para o ano de 2011, para nenhum dos 4 países listados neste estudo. Logo, para fim deste exercício, foi inserido o valor da média dos anos 2010 e 2012 de cada país em sua respectiva célula.

As três séries de dados podem ser consideradas satisfatórias para a realização desta pesquisa, uma vez que todas possuem dados de mais de 20 anos, sem interrupções.

4.2.1 As Variáveis

São 4 as variáveis desta pesquisa, incluindo o ano a que se referem os dados. Estas variáveis são melhor explicadas na Tabela 4.1.

Tabela 4.1. As Variáveis

Variável	Significado	Tipo	Unidade de Medida
Ano	É o ano a que se refere o dado de cada variável e de cada país.	Variável Categórica	Formato: AAAA
TerMarProt_XX <i>Terrestrial and marine protected areas</i> (áreas terrestres e marinhas sob proteção) (unit: % of total territorial area) Período de 1990 a 2012	É a taxa que representa as áreas terrestres e marinhas que estão sob proteção em determinado país. É medida considerando a relação entre o total da área protegida pelo total da área territorial do país, em determinado ano. O sufixo XXX, no nome da variável, denota o nome abreviado do país em análise, por exemplo: BRA para Brasil, MEX para México, POR para Portugal e SPA para Espanha.	Variável Quantitativa	Percentual (%)
Forest_XXX <i>Forest area</i> (área florestal) (unit: sq. km) Período de 1990 a 2013	Refere-se a medida da área coberta por florestas que ocupam acima de 0.5 hectares, com árvores maiores que 5 metros de altura. Não inclui áreas destinadas a agricultura e zonas urbanas. É caracterizada pela presença de árvores e pela não utilização destas terras para qualquer outra finalidade. O sufixo XXX, no nome da variável, denota o nome abreviado do país em análise, por exemplo: BRA para Brasil, MEX para México, POR para Portugal e SPA para Espanha.	Variável Quantitativa	Quilômetros quadrados (km ²)
Pop_XXX <i>Population</i> (população total de um país) (unit: Persons) Período de 1990 a 2013	Mede a quantidade total de indivíduos que vivem no país (residentes), independente da nacionalidade, exceto refugiados e asilados. O sufixo XXX, no nome da variável, denota o nome abreviado do país em análise, por exemplo: BRA para Brasil, MEX para México, POR para Portugal e SPA para Espanha.	Variável Quantitativa	Quantidade de indivíduos

4.3. ANÁLISE DAS VARIÁVEIS

Cada variável será analisada utilizando gráficos para demonstrar o comportamento histórico da série, linhas de tendência, funções, erros das funções, além de extrapolações estatísticas.

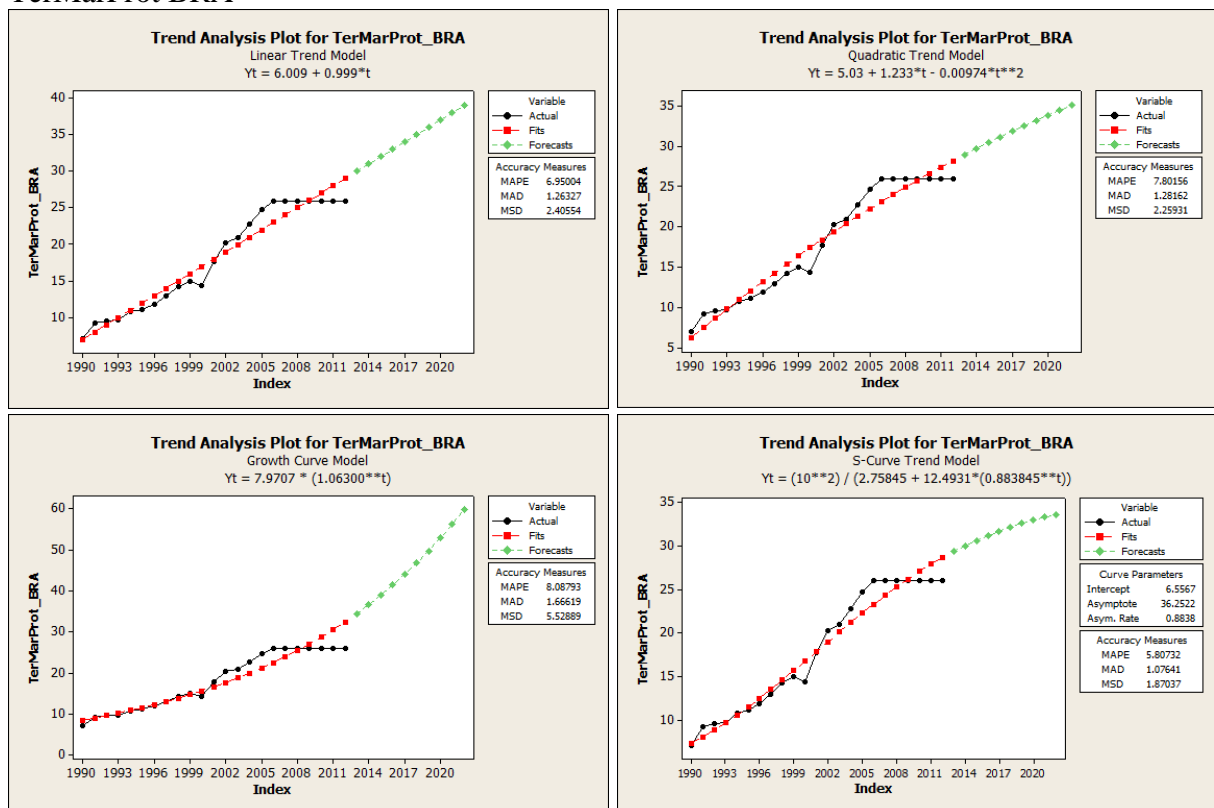
Na seção que segue, analisaremos a tendência da série de dados, buscando encontrar a função (linear, quadrática, exponencial ou s-curve) que melhor a represente. Para medir a eficiência da função ou da curva de tendência, analisaremos os erros: MAPE (*Mean Average Percentual Error*), MAD (*Mean Absolute Deaviation*) e MSD (*Mean Standard Deviation*).

4.3.1 Variável: TerMarProt_BRA - Áreas terrestres e marinhas sob proteção, no Brasil

4.3.1.1. A análise de tendência

Segue abaixo gráficos incluindo a função, a linha de tendência que representa cada função e os erros mencionados acima.

TerMarProt BRA



Para podermos visualizar qual a melhor função para representar a série de dados acima, podemos analisar a tabela abaixo, que compara os três tipos de erros para cada uma das quatro funções:

Estatística	Linear	Quadrática	Exponencial	S-Curve
MAPE	6.95004	7.80156	8.08793	5.80732
MAD	1.26327	1.28162	1.66619	1.07641
MSD	2.40554	2.25931	5.52889	1.87037

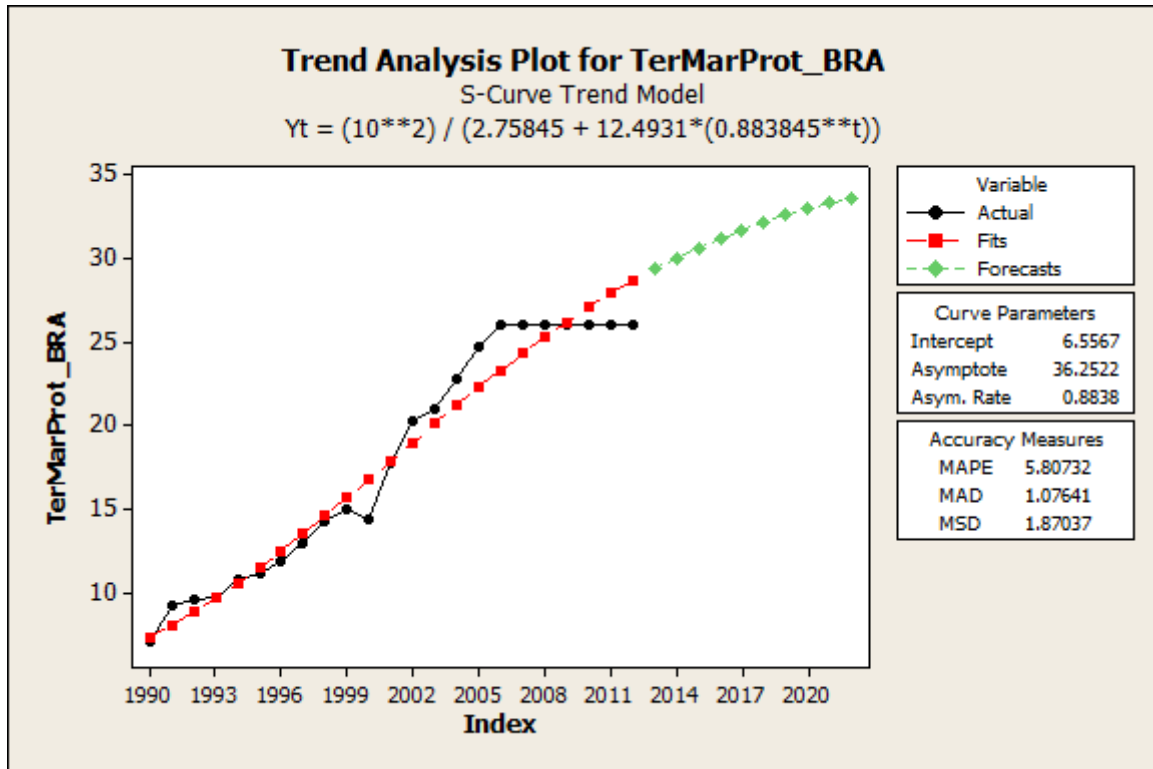
Percebemos que o modelo **S-Curve** é o que melhor se adapta à nossa série de dados. Assim, esta será a função utilizada para as projeções (*forecasts*).

4.3.1.2. As previsões

Utilizando a equação da função **S-Curve** apresentada no gráfico acima, temos os seguintes valores previstos (*forecasts*) para os próximos 10 anos em relação às áreas terrestres e marinhas sob proteção, no Brasil.

Year	Forecast
2013	29.3802
2014	30.0416
2015	30.6516
2016	31.2117
2017	31.7240
2018	32.1911
2019	32.6155
2020	33.0000
2021	33.3475
2022	33.6608

Os valores acima podem ser melhor ilustrados de forma gráfica, conforme abaixo:



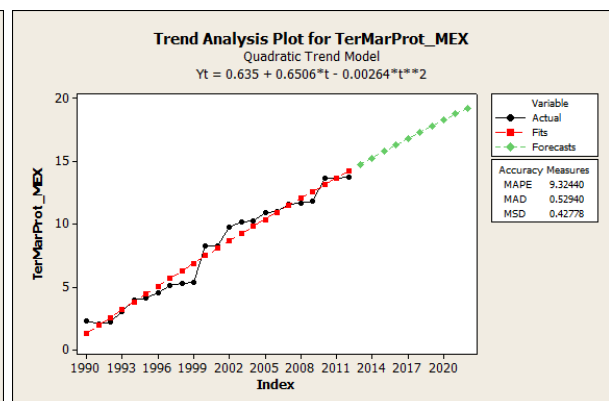
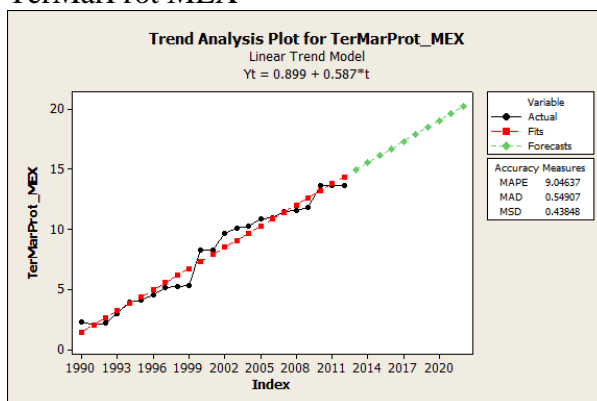
Ressaltamos que trata-se de uma previsão efetuada somente com base nos valores históricos da própria variável. Certamente esta metodologia apresenta deficiências ao assumir que o comportamento passado se repetirá. Do mesmo modo, talvez a utilização de regressões múltiplas poderiam ser mais eficientes neste trabalho de previsão.

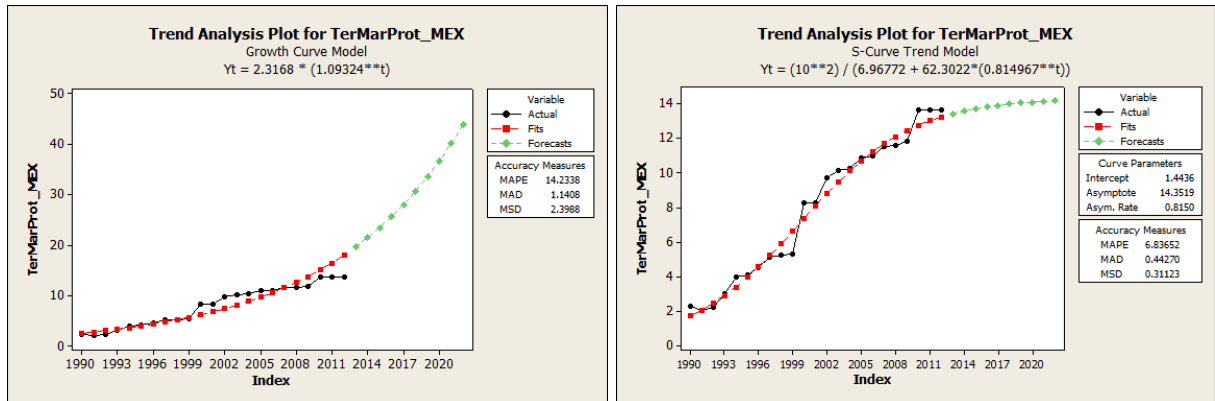
4.3.2 Variável: TerMarProt_MEX - Áreas terrestres e marinhas sob proteção, no México

4.3.2.1. A análise de tendência

Segue abaixo gráficos incluindo a função, a linha de tendência que representa cada função e os erros mencionados acima.

TerMarProt MEX





Para podermos visualizar qual a melhor função para representar a série de dados acima, podemos analisar a tabela abaixo, que compara os três tipos de erros para cada uma das quatro funções:

Estadística	Linear	Quadrática	Exponencial	S-Curve
MAPE	9.04637	9.32440	14.2338	6.83652
MAD	0.54907	0.52940	1.1408	0.44270
MSD	0.43848	0.42778	2.3988	0.31123

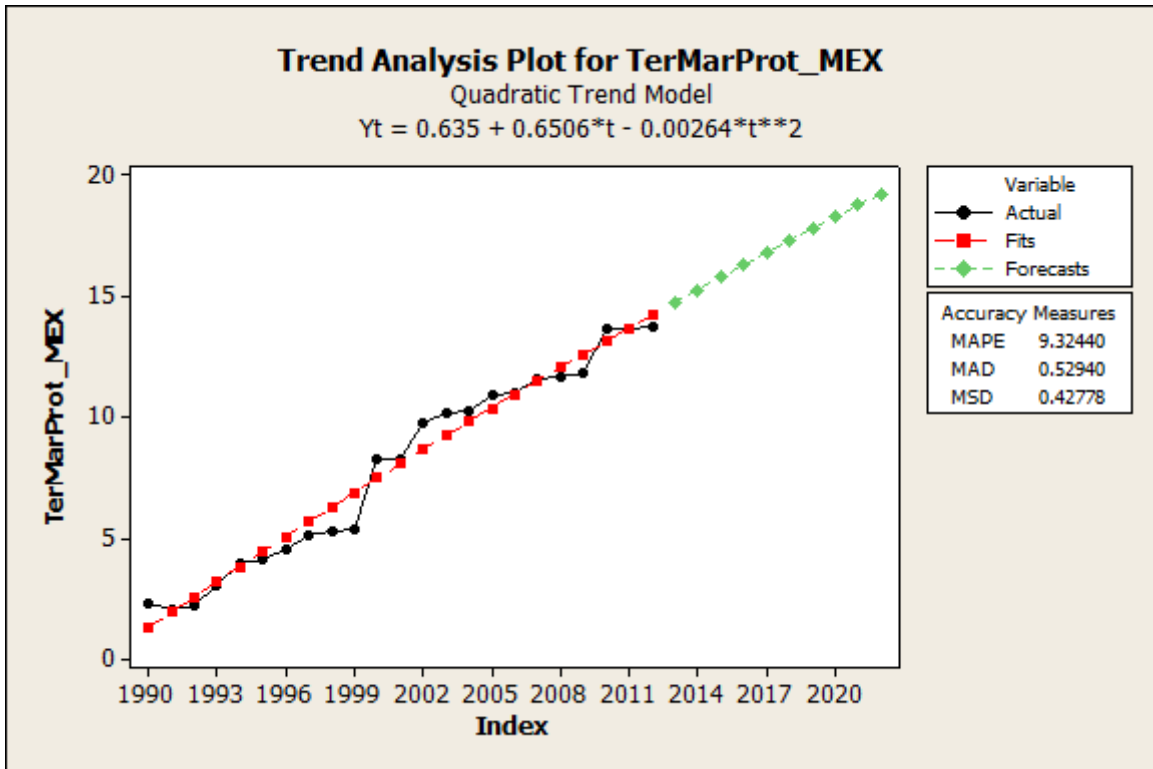
Notamos que o modelo **Quadrática** é o que melhor se adapta à nossa série de dados. Assim, esta será a função utilizada para as projeções (*forecasts*).

4.3.2.2. As previsões

Utilizando a equação da função **Quadrática** apresentada no gráfico acima, temos os seguintes valores previstos (*forecasts*) para os próximos 10 anos em relação às áreas terrestres e marinhas sob proteção, no México.

Year	Forecast
2013	14.7313
2014	15.2528
2015	15.7690
2016	16.2799
2017	16.7855
2018	17.2858
2019	17.7809
2020	18.2707
2021	18.7553
2022	19.2346

Os valores acima podem ser melhor ilustrados de forma gráfica, conforme abaixo:



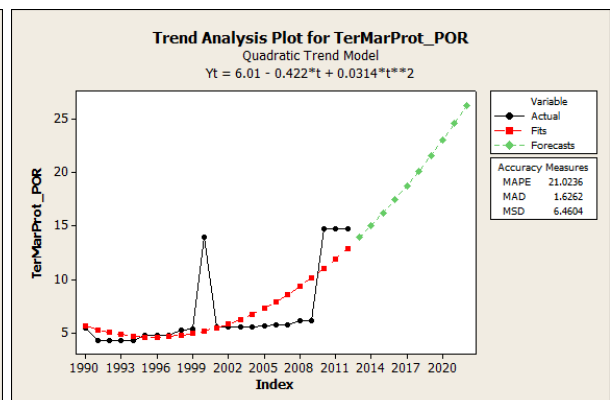
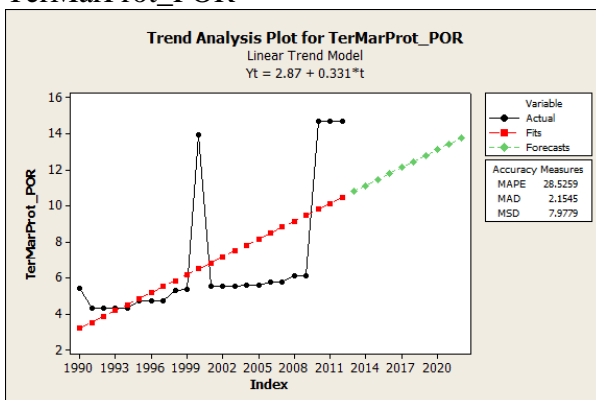
No gráfico acima, trata-se de uma previsão efetuada somente com base nos valores históricos da própria variável. Certamente esta metodologia apresenta deficiências ao assumir que o comportamento passado se repetirá. Do mesmo modo, talvez a utilização de regressões múltiplas poderiam ser mais eficientes neste trabalho de previsão.

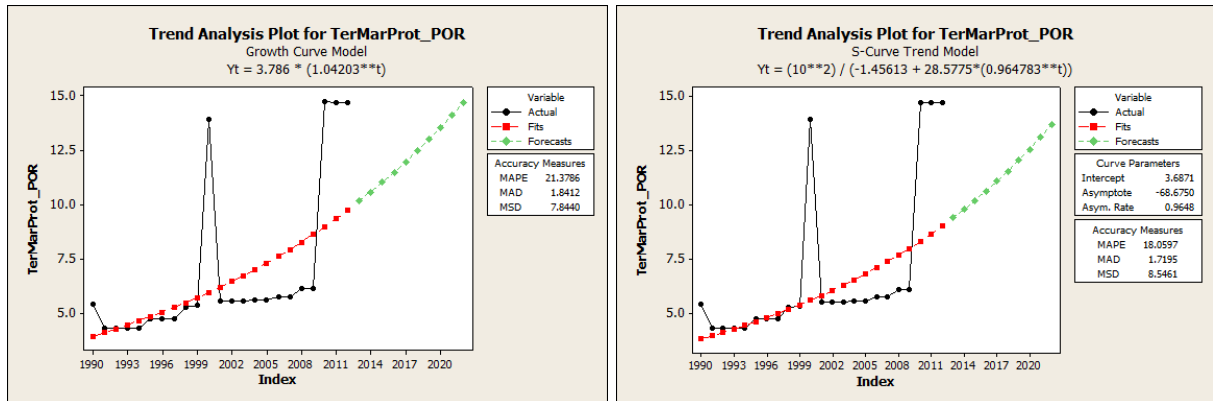
4.3.3 Variável: TerMarProt_POR - Áreas terrestres e marinhas sob proteção, em Portugal

4.3.3.1. A análise de tendência

Segue abaixo gráficos incluindo a função, a linha de tendência que representa cada função e os erros mencionados acima.

TerMarProt_POR





Para podermos visualizar qual a melhor função para representar a série de dados acima, podemos analisar a tabela abaixo, que compara os três tipos de erros para cada uma das quatro funções:

Estadística	Linear	Quadrática	Exponencial	S-Curve
MAPE	28.5259	21.0236	21.3786	18.0597
MAD	2.1545	1.6262	1.8412	1.7195
MSD	7.9779	6.4604	7.8440	8.5461

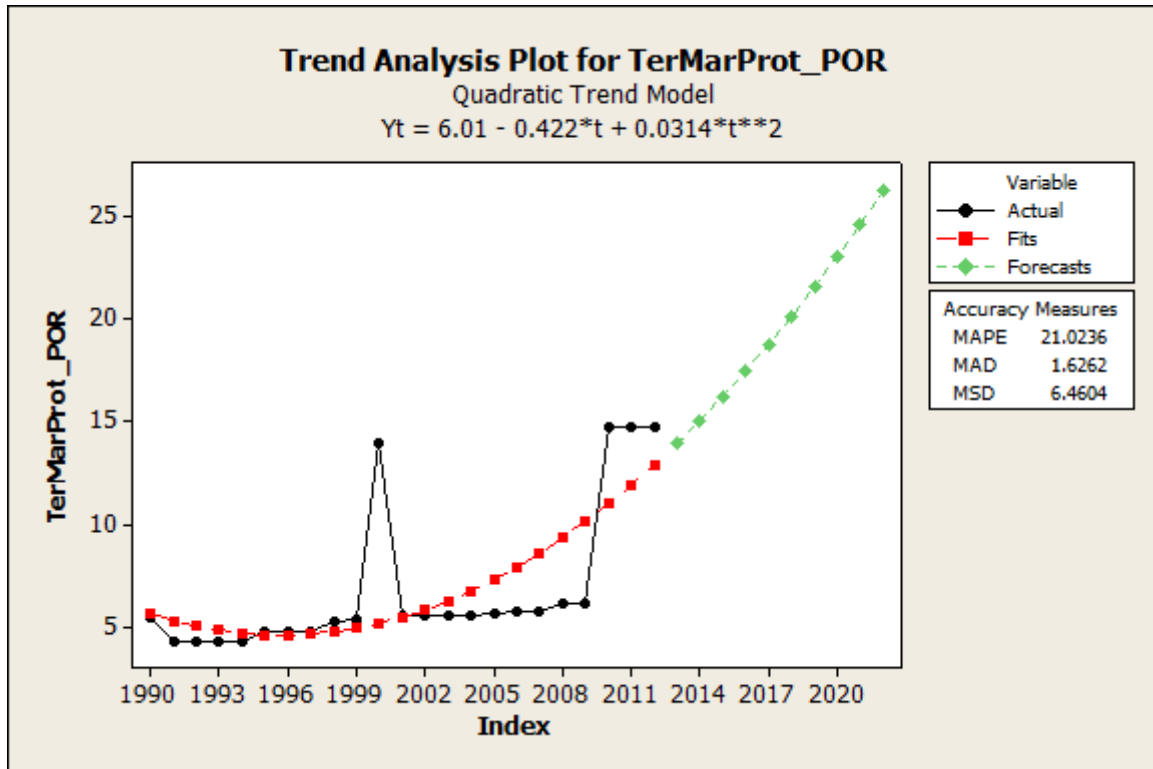
Percebemos que o modelo **Quadrática** é o que melhor se adapta à nossa série de dados. Assim, esta será a função utilizada para as projeções (*forecasts*).

4.3.3.2. As previsões

Utilizando a equação da função **Quadrática** apresentada no gráfico acima, temos os seguintes valores previstos (*forecasts*) para os próximos 10 anos em relação às áreas terrestres e marinhas sob proteção, em Portugal.

Year	Forecast
2013	13.9527
2014	15.0684
2015	16.2469
2016	17.4882
2017	18.7922
2018	20.1591
2019	21.5887
2020	23.0811
2021	24.6363
2022	26.2543

Os valores acima podem ser melhor ilustrados de forma gráfica, conforme abaixo:



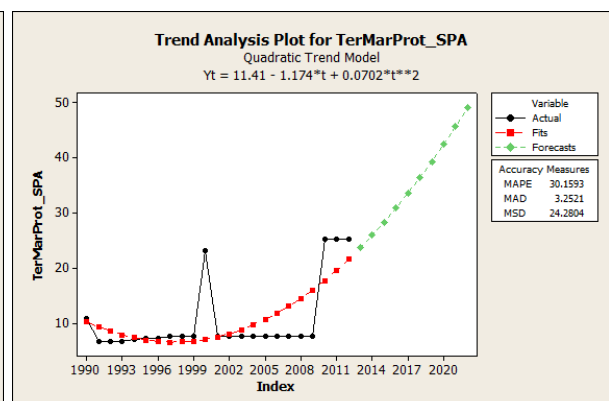
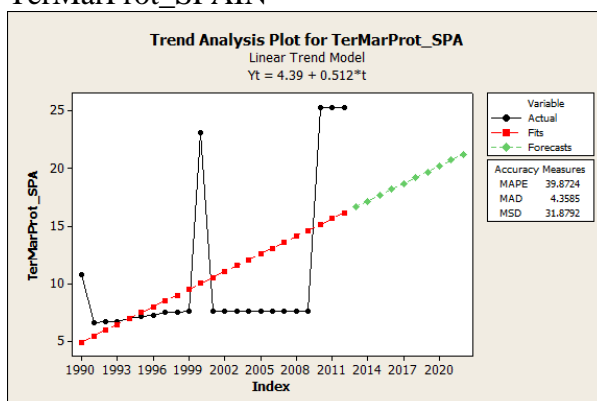
Ressaltamos que trata-se de uma previsão efetuada somente com base nos valores históricos da própria variável. Certamente esta metodologia apresenta deficiências ao assumir que o comportamento passado se repetirá. Do mesmo modo, talvez a utilização de regressões múltiplas poderiam ser mais eficientes neste trabalho de previsão.

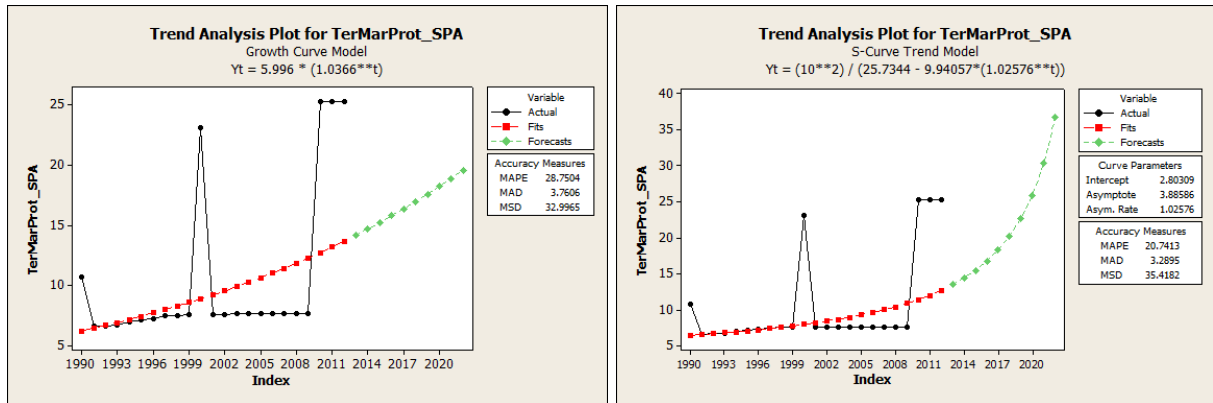
4.3.4 Variável: TerMarProt_SPA - Áreas terrestres e marinhas sob proteção, na Espanha

4.3.4.1. A análise de tendência

Segue abaixo gráficos incluindo a função, a linha de tendência que representa cada função e os erros mencionados acima.

TerMarProt_SPAIN





Para podermos visualizar qual a melhor função para representar a série de dados acima, podemos analisar a tabela abaixo, que compara os três tipos de erros para cada uma das quatro funções:

Estadística	Linear	Quadrática	Exponencial	S-Curve
MAPE	39.8724	30.1593	28.7504	20.7413
MAD	4.3585	3.2521	3.7606	3.2895
MSD	31.8792	24.2804	32.9965	35.4182

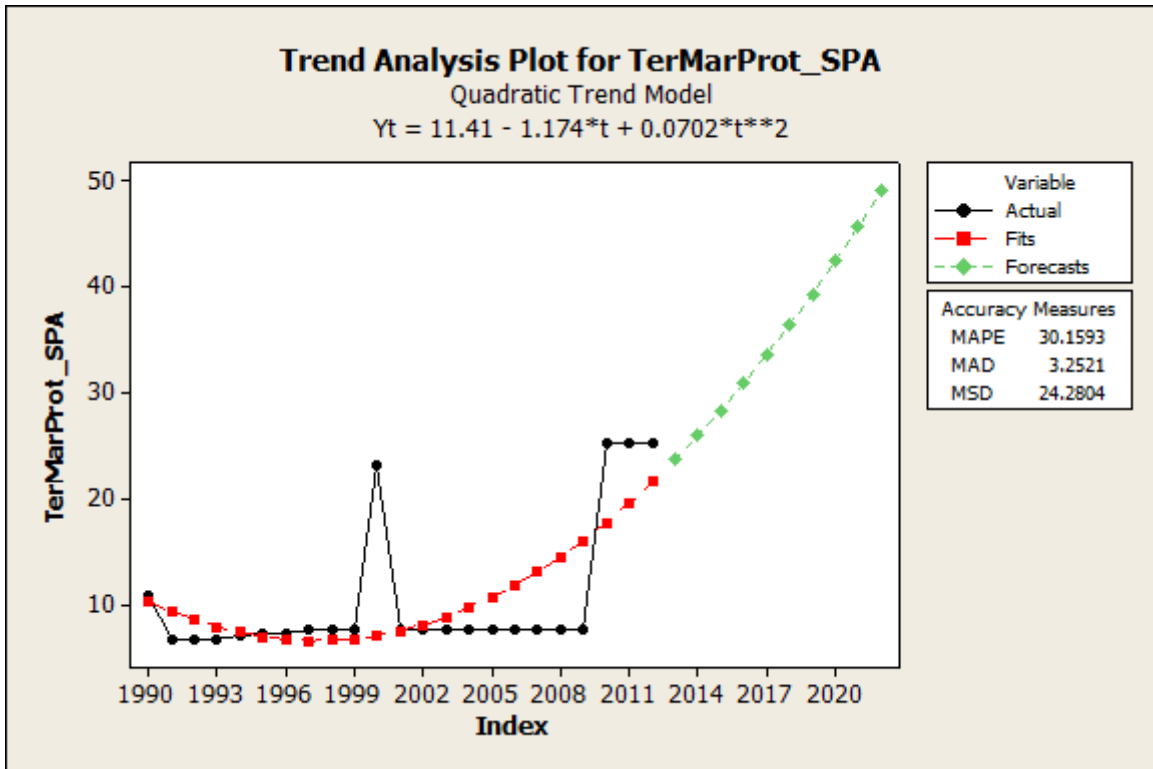
Notamos que o modelo **Quadrática** é o que melhor se adapta à nossa série de dados. Assim, esta será a função utilizada para as projeções (*forecasts*).

4.3.4.2. As previsões

Utilizando a equação da função **Quadrática** apresentada no gráfico acima, temos os seguintes valores previstos (*forecasts*) para os próximos 10 anos em relação às áreas terrestres e marinhas sob proteção, na Espanha.

Year	Forecast
2013	23.7051
2014	25.9735
2015	28.3825
2016	30.9319
2017	33.6218
2018	36.4522
2019	39.4231
2020	42.5344
2021	45.7863
2022	49.1786

Os valores acima podem ser melhor ilustrados de forma gráfica, conforme abaixo:



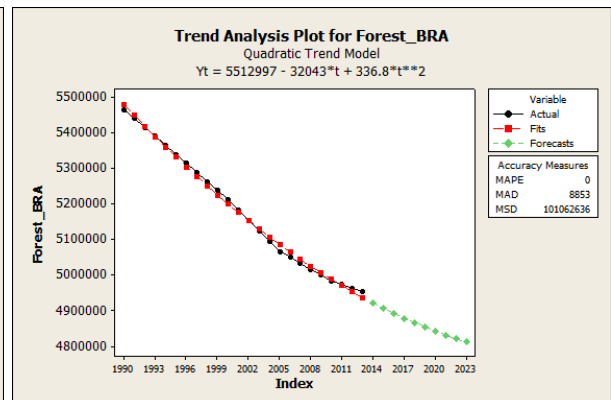
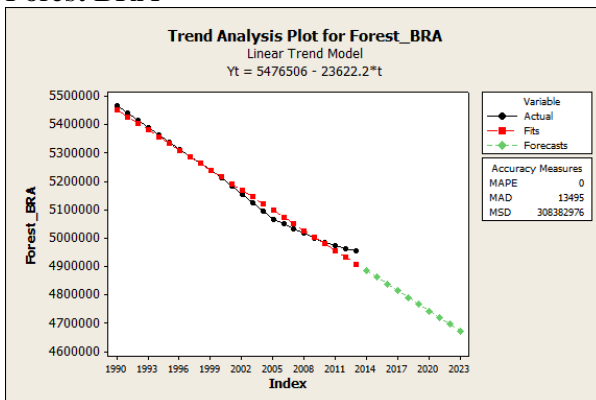
No gráfico acima, trata-se de uma previsão efetuada somente com base nos valores históricos da própria variável. Certamente esta metodologia apresenta deficiências ao assumir que o comportamento passado se repetirá. Do mesmo modo, talvez a utilização de regressões múltiplas poderiam ser mais eficientes neste trabalho de previsão.

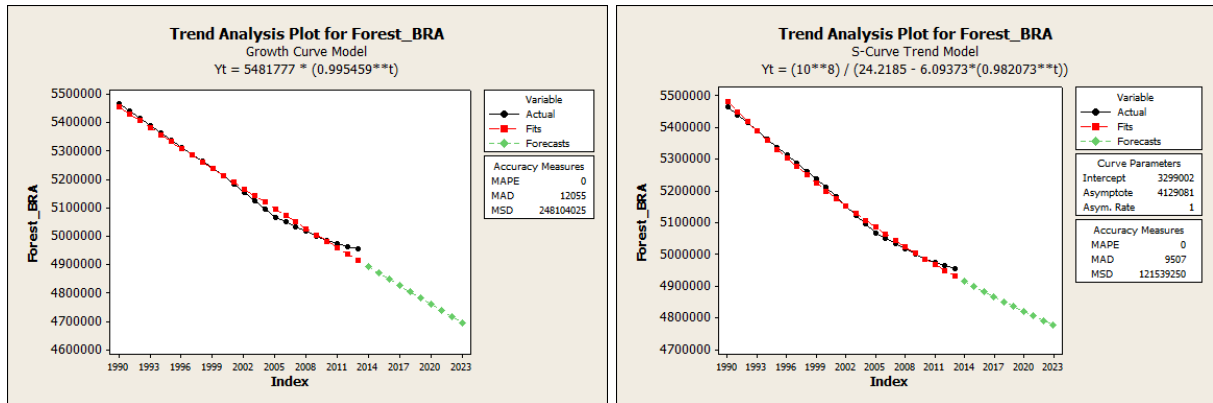
4.3.5 Variável: Forest_BRA - Áreas de florestas, no Brasil

4.3.5.1. A análise de tendência

Segue abaixo gráficos incluindo a função, a linha de tendência que representa cada função e os erros mencionados acima.

Forest BRA





Para podermos visualizar qual a melhor função para representar a série de dados acima, podemos analisar a tabela abaixo, que compara os três tipos de erros para cada uma das quatro funções:

Estatística	Linear	Quadrática	Exponencial	S-Curve
MAPE	0	0	0	0
MAD	13495	8853	12055	9507
MSD	308382976	101062636	248104025	121539250

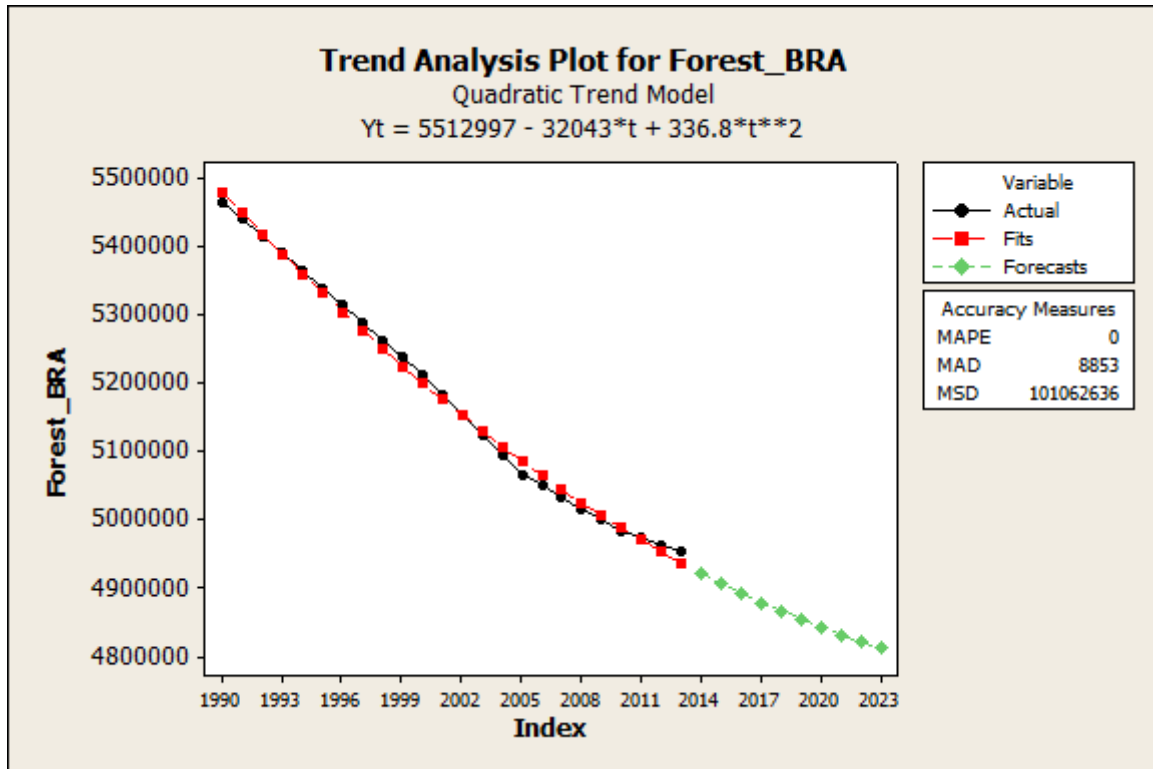
Percebemos que o modelo **Quadrática** é o que melhor se adapta à nossa série de dados. Assim, esta será a função utilizada para as projeções (*forecasts*).

4.3.5.2. As previsões

Utilizando a equação da função **Quadrática** apresentada no gráfico acima, temos os seguintes valores previstos (*forecasts*) para os próximos 10 anos em relação às áreas de florestas sob proteção, no Brasil.

Year	Forecast
2014	4922441
2015	4907577
2016	4893386
2017	4879869
2018	4867026
2019	4854856
2020	4843360
2021	4832538
2022	4822389
2023	4812914

Os valores acima podem ser melhor ilustrados de forma gráfica, conforme abaixo:



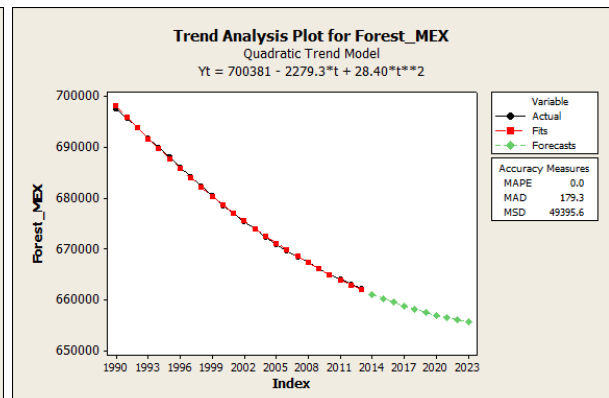
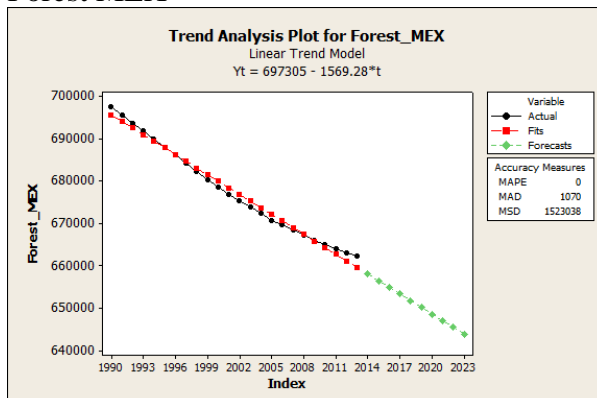
Ressaltamos que trata-se de uma previsão efetuada somente com base nos valores históricos da própria variável. Certamente esta metodologia apresenta deficiências ao assumir que o comportamento passado se repetirá. Do mesmo modo, talvez a utilização de regressões múltiplas poderiam ser mais eficientes neste trabalho de previsão.

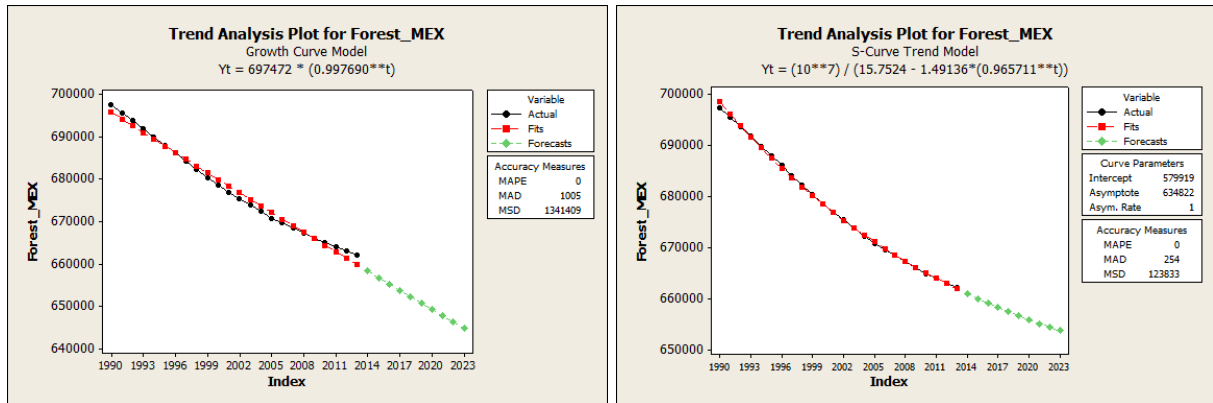
4.3.6 Variável: Forest_MEX - Áreas de florestas, no México

4.3.6.1. A análise de tendência

Segue abaixo gráficos incluindo a função, a linha de tendência que representa cada função e os erros mencionados acima.

Forest MEX





Para podermos visualizar qual a melhor função para representar a série de dados acima, podemos analisar a tabela abaixo, que compara os três tipos de erros para cada uma das quatro funções:

Estadística	Linear	Quadrática	Exponencial	S-Curve
MAPE	0	0.0	0	0
MAD	1070	179.3	1005	254
MSD	1523038	49395.6	1341409	123833

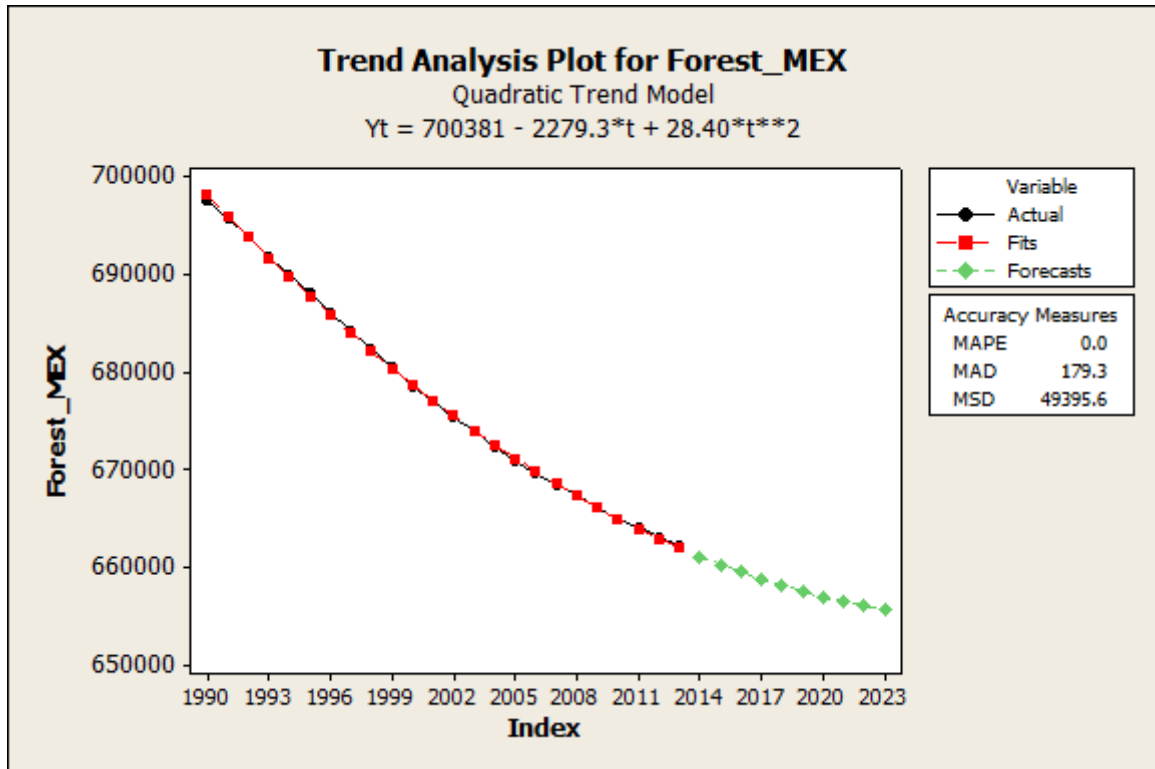
Notamos que o modelo **Quadrática** é o que melhor se adapta à nossa série de dados. Assim, esta será a função utilizada para as projeções (*forecasts*).

4.3.6.2. As previsões

Utilizando a equação da função **Quadrática** apresentada no gráfico acima, temos os seguintes valores previstos (*forecasts*) para os próximos 10 anos em relação às áreas de florestas, no México.

Year	Forecast
2014	661149
2015	660319
2016	659544
2017	658827
2018	658167
2019	657563
2020	657016
2021	656526
2022	656092
2023	655716

Os valores acima podem ser melhor ilustrados de forma gráfica, conforme abaixo:



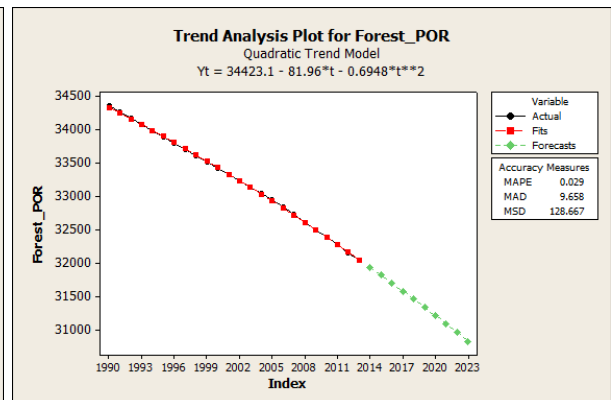
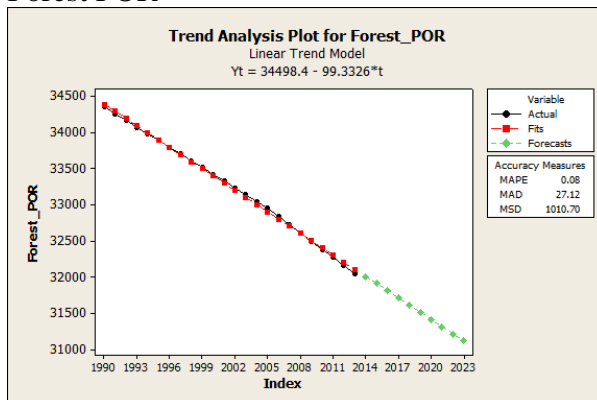
No gráfico acima, trata-se de uma previsão efetuada somente com base nos valores históricos da própria variável. Certamente esta metodologia apresenta deficiências ao assumir que o comportamento passado se repetirá. Do mesmo modo, talvez a utilização de regressões múltiplas poderiam ser mais eficientes neste trabalho de previsão.

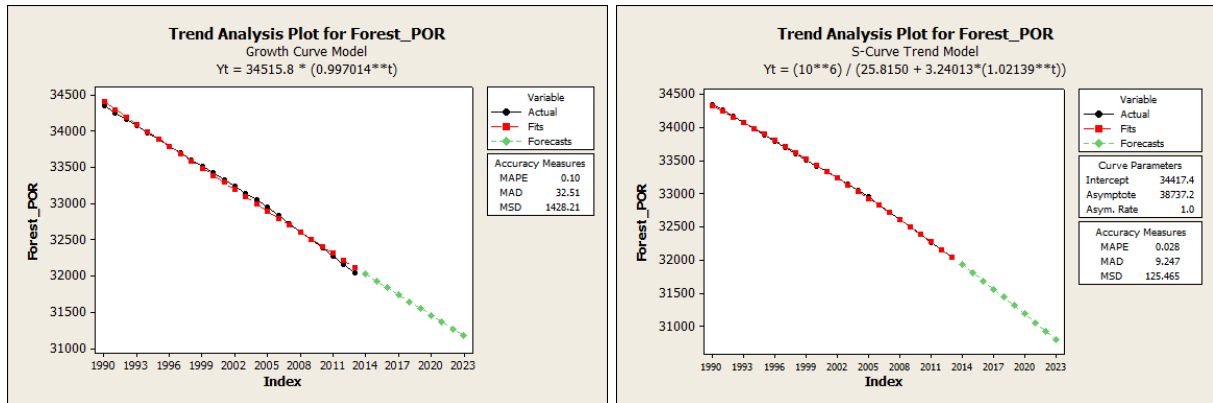
4.3.7 Variável: Forest_POR - Áreas de florestas, em Portugal

4.3.7.1. A análise de tendência

Segue abaixo gráficos incluindo a função, a linha de tendência que representa cada função e os erros mencionados acima.

Forest POR





Para podermos visualizar qual a melhor função para representar a série de dados acima, podemos analisar a tabela abaixo, que compara os três tipos de erros para cada uma das quatro funções:

Estadística	Linear	Quadrática	Exponencial	S-Curve
MAPE	0.08	0.029	0.10	0.028
MAD	27.12	9.658	32.51	9.247
MSD	1010.70	128.667	1428.21	125.465

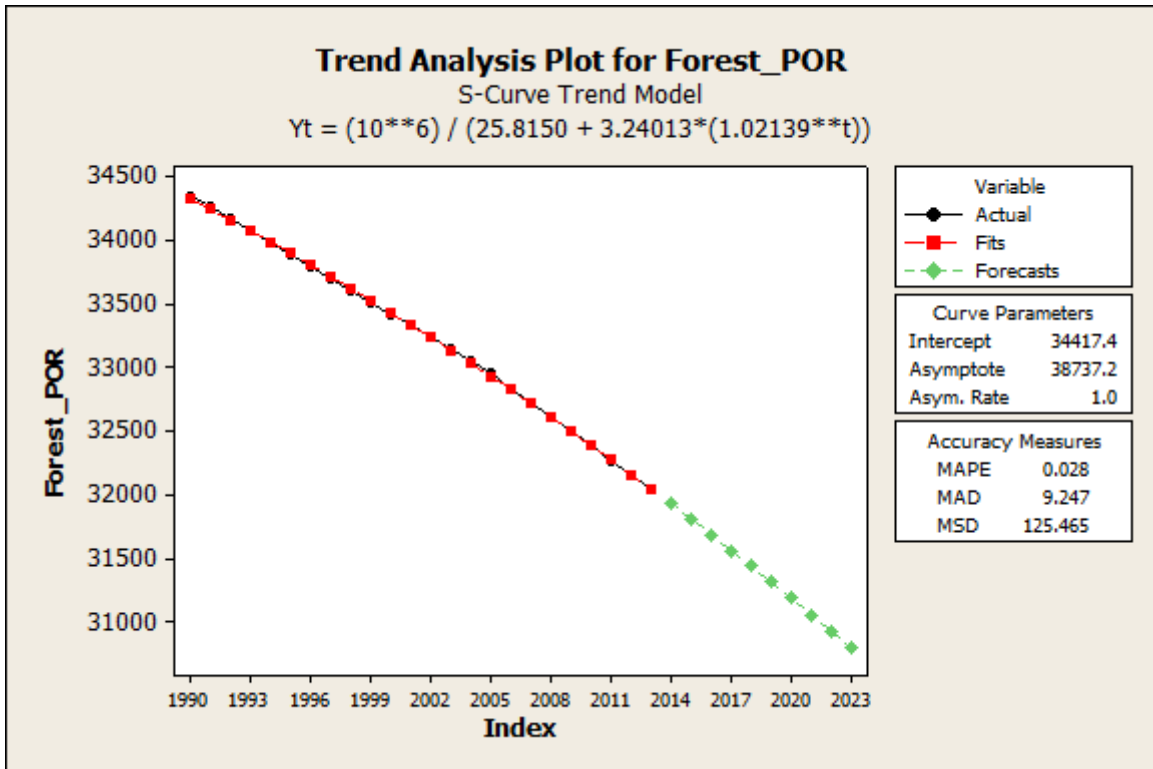
Percebemos que o modelo **S-Curve** é o que melhor se adapta à nossa série de dados. Assim, esta será a função utilizada para as projeções (*forecasts*).

4.3.7.2. As previsões

Utilizando a equação da função **S-Curve** apresentada no gráfico acima, temos os seguintes valores previstos (*forecasts*) para os próximos 10 anos em relação às áreas de florestas sob proteção, em Portugal.

<u>Year</u>	<u>Forecast</u>
2014	31933.2
2015	31813.6
2016	31692.4
2017	31569.6
2018	31445.1
2019	31319.0
2020	31191.2
2021	31061.8
2022	30930.6
2023	30797.9

Os valores acima podem ser melhor ilustrados de forma gráfica, conforme abaixo:



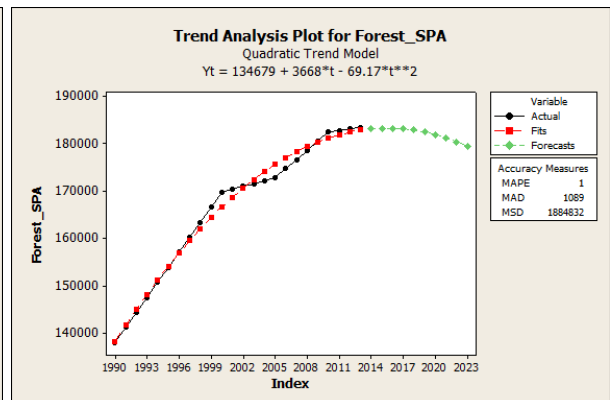
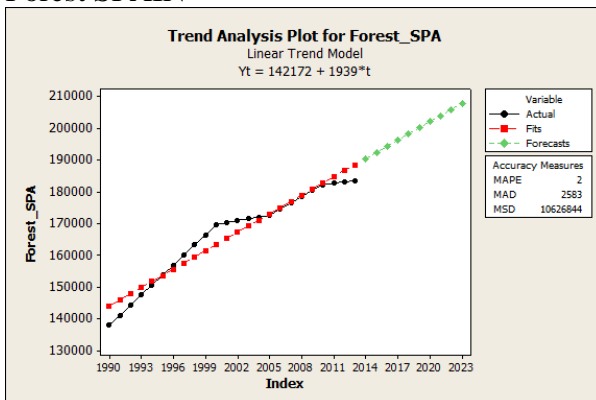
Ressaltamos que trata-se de uma previsão efetuada somente com base nos valores históricos da própria variável. Certamente esta metodologia apresenta deficiências ao assumir que o comportamento passado se repetirá. Do mesmo modo, talvez a utilização de regressões múltiplas poderiam ser mais eficientes neste trabalho de previsão.

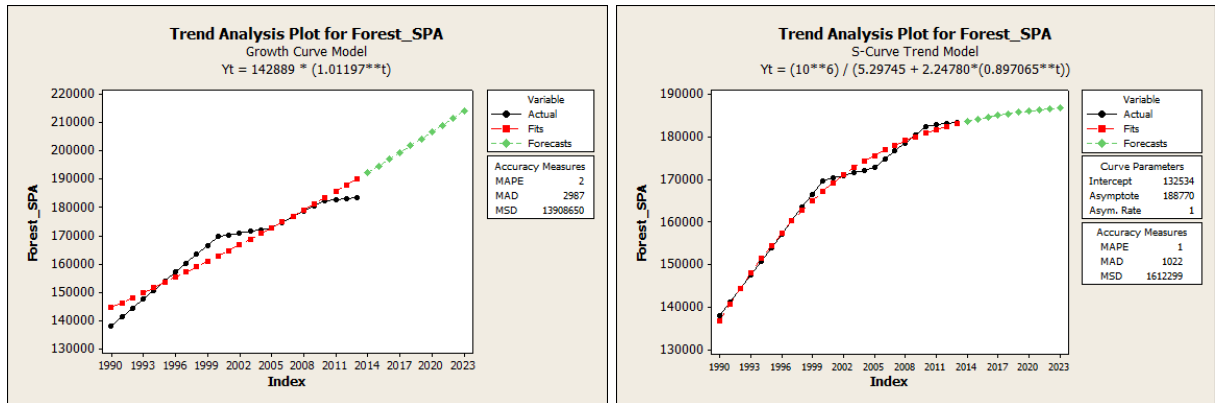
4.3.8 Variável: Forest_SPA - Áreas de florestas, na Espanha

4.3.8.1. A análise de tendência

Segue abaixo gráficos incluindo a função, a linha de tendência que representa cada função e os erros mencionados acima.

Forest SPAIN





Para podermos visualizar qual a melhor função para representar a série de dados acima, podemos analisar a tabela abaixo, que compara os três tipos de erros para cada uma das quatro funções:

Estadística	Linear	Quadrática	Exponencial	S-Curve
MAPE	2	1	2	1
MAD	2583	1089	2987	1022
MSD	10626844	1884832	13908650	1612299

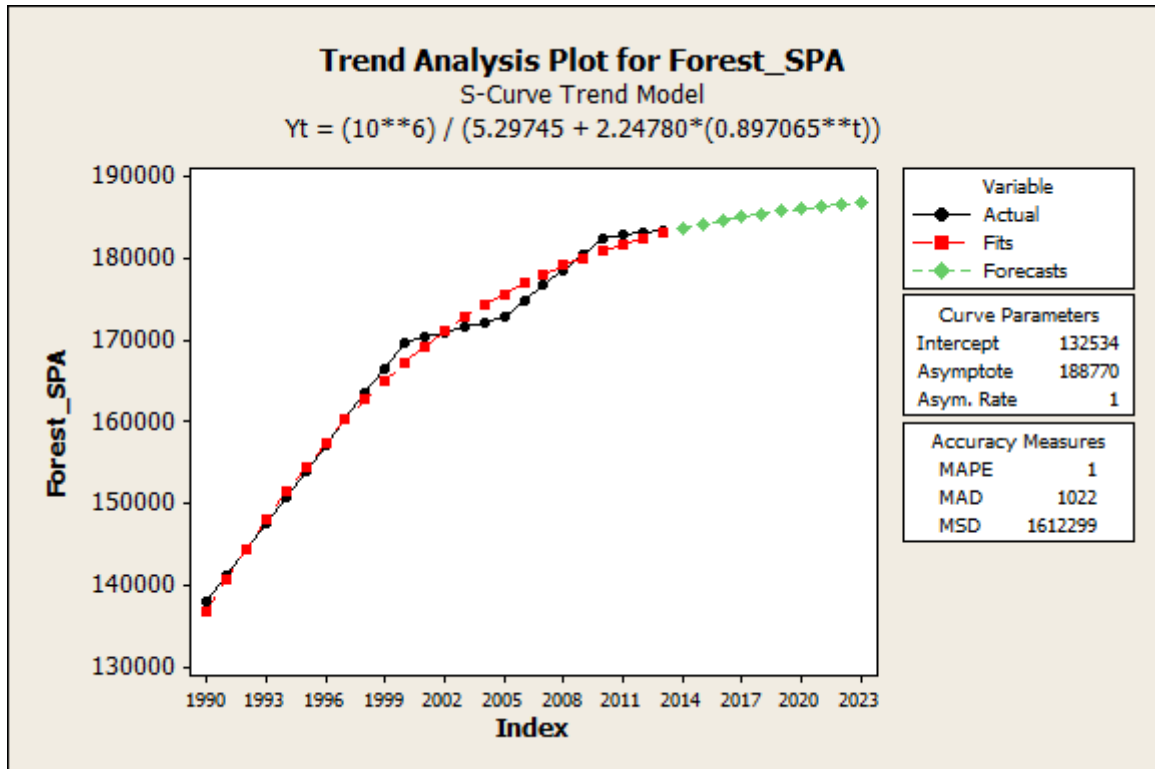
Notamos que o modelo **S-Curve** é o que melhor se adapta à nossa série de dados. Assim, esta será a função utilizada para as projeções (*forecasts*).

4.3.8.2. As previsões

Utilizando a equação da função **S-Curve** apresentada no gráfico acima, temos os seguintes valores previstos (*forecasts*) para os próximos 10 anos em relação às áreas de florestas, na Espanha.

<u>Year</u>	<u>Forecast</u>
2014	183615
2015	184133
2016	184600
2017	185021
2018	185400
2019	185741
2020	186048
2021	186325
2022	186574
2023	186797

Os valores acima podem ser melhor ilustrados de forma gráfica, conforme abaixo:



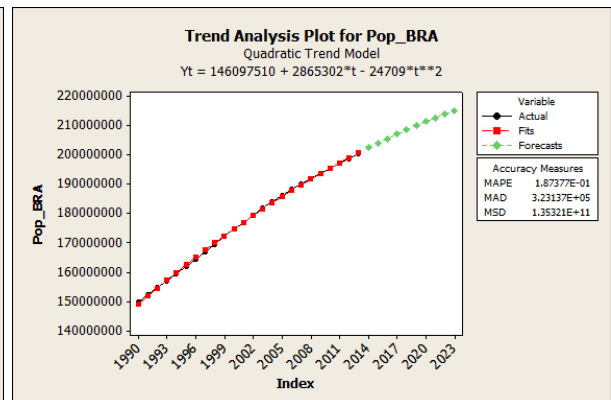
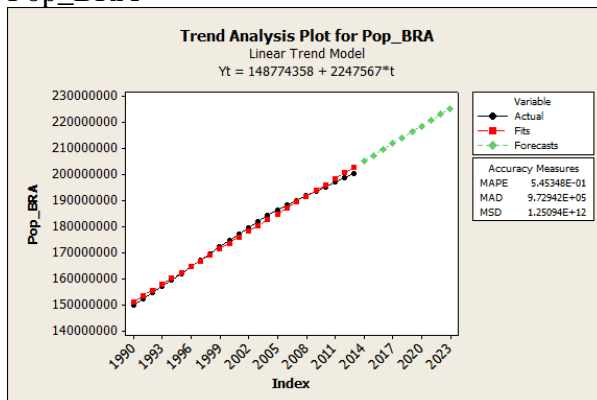
No gráfico acima, trata-se de uma previsão efetuada somente com base nos valores históricos da própria variável. Certamente esta metodologia apresenta deficiências ao assumir que o comportamento passado se repetirá. Do mesmo modo, talvez a utilização de regressões múltiplas poderiam ser mais eficientes neste trabalho de previsão.

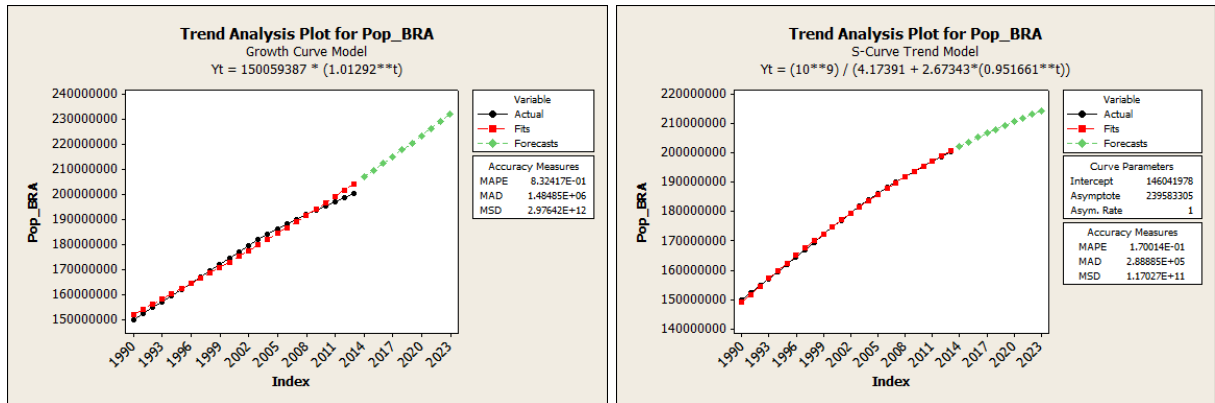
4.3.9 Variável: Pop_POR – População Total no Brasil

4.3.9.1. A análise de tendência

Segue abaixo gráficos incluindo a função, a linha de tendência que representa cada função e os erros mencionados acima.

Pop_BRA





Para podermos visualizar qual a melhor função para representar a série de dados acima, podemos analisar a tabela abaixo, que compara os três tipos de erros para cada uma das quatro funções:

Estadística	Linear	Quadrática	Exponencial	S-Curve
MAPE	5.45348E-01	1.87377E-01	8.32417E-01	1.70014E-01
MAD	9.72942E+05	3.23137E+05	1.48485E+06	2.88885E+05
MSD	1.25094E+12	1.35321E+11	2.97642E+12	1.17027E+11

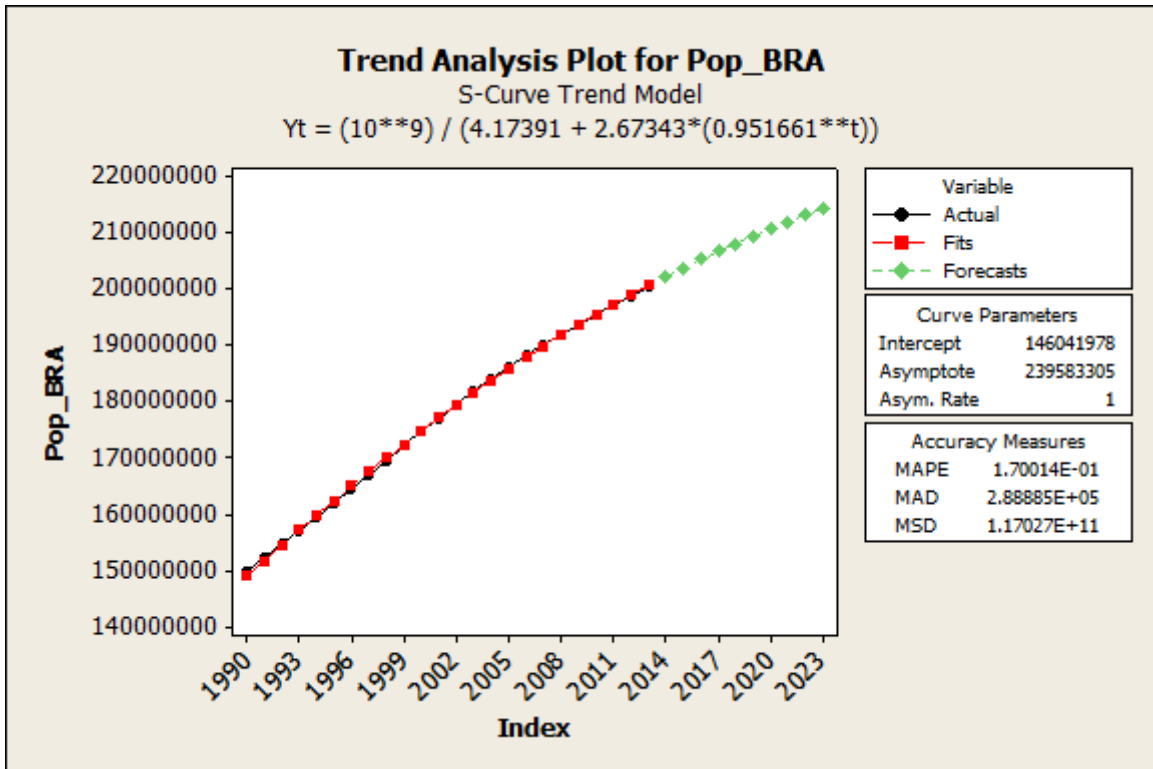
Percebemos que o modelo **S-Curve** é o que melhor se adapta à nossa série de dados. Assim, esta será a função utilizada para as projeções (*forecasts*).

4.3.9.2. As previsões

Utilizando a equação da função **S-Curve** apresentada no gráfico acima, temos os seguintes valores previstos (*forecasts*) para os próximos 10 anos em relação população total, no Brasil.

Year	Forecast
2014	202077154
2015	203617994
2016	205106333
2017	206543075
2018	207929185
2019	209265680
2020	210553623
2021	211794119
2022	212988301
2023	214137331

Os valores acima podem ser melhor ilustrados de forma gráfica, conforme abaixo:



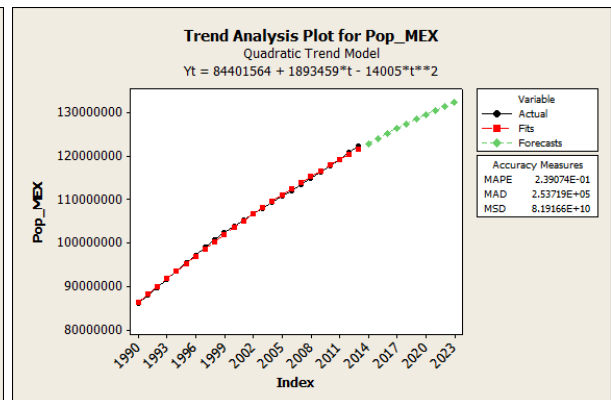
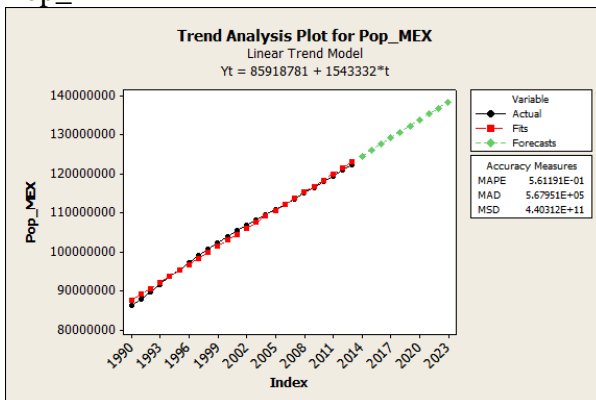
Ressaltamos que trata-se de uma previsão efetuada somente com base nos valores históricos da própria variável. Certamente esta metodologia apresenta deficiências ao assumir que o comportamento passado se repetirá. Do mesmo modo, talvez a utilização de regressões múltiplas poderiam ser mais eficientes neste trabalho de previsão.

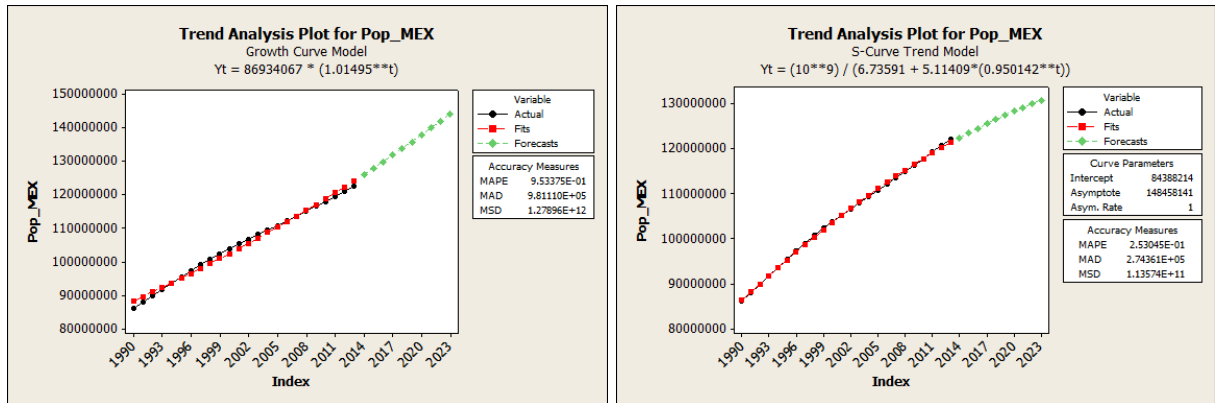
4.3.10 Variável: Pop_MEX – População Total no México

4.3.10.1. A análise de tendência

Segue abaixo gráficos incluindo a função, a linha de tendência que representa cada função e os erros mencionados acima.

Pop_MEX





Para podermos visualizar qual a melhor função para representar a série de dados acima, podemos analisar a tabela abaixo, que compara os três tipos de erros para cada uma das quatro funções:

Estadística	Linear	Quadrática	Exponencial	S-Curve
MAPE	5.61191E-01	2.39074E-01	9.53375E-01	2.53045E-01
MAD	5.67951E+05	2.53719E+05	9.81110E+05	2.74361E+05
MSD	4.40312E+11	8.19166E+10	1.27896E+12	1.13574E+11

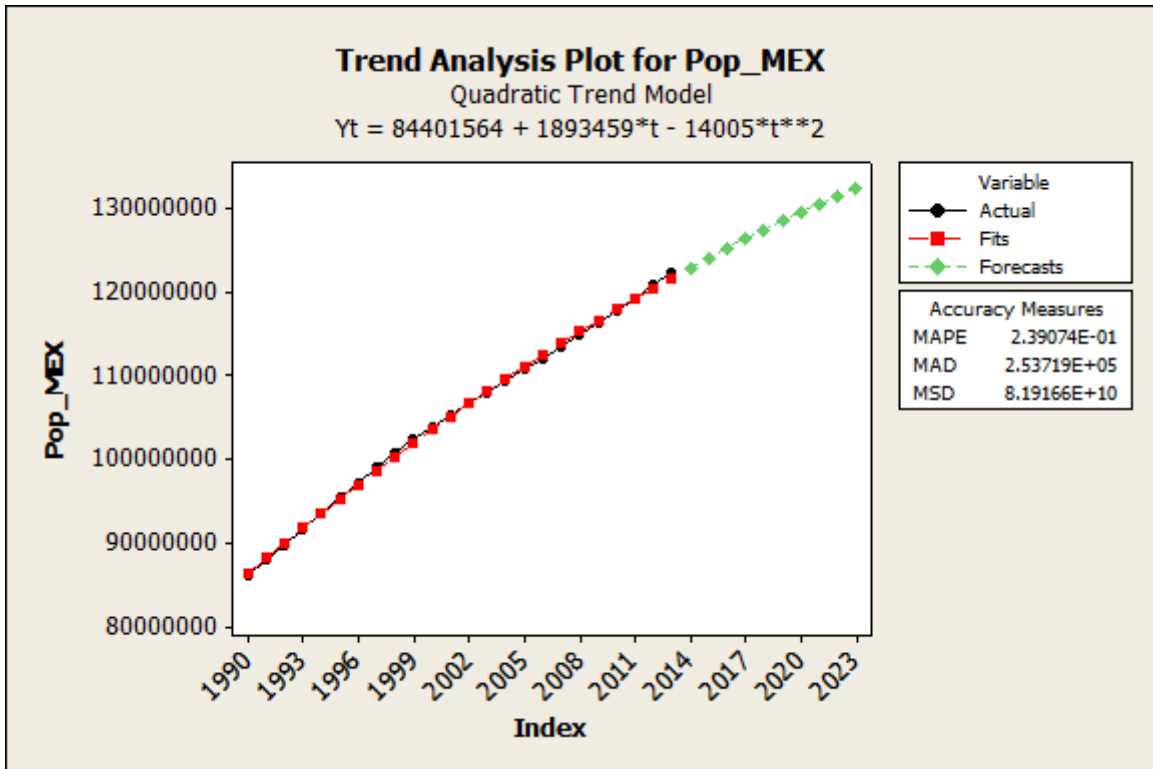
Notamos que o modelo **Quadrática** é o que melhor se adapta à nossa série de dados. Assim, esta será a função utilizada para as projeções (*forecasts*).

4.3.10.2. As previsões

Utilizando a equação da função **Quadrática** apresentada no gráfico acima, temos os seguintes valores previstos (*forecasts*) para os próximos 10 anos em relação a população total no México.

Year	Forecast
2014	122984874
2015	124164074
2016	125315265
2017	126438445
2018	127533615
2019	128600774
2020	129639924
2021	130651063
2022	131634193
2023	132589312

Os valores acima podem ser melhor ilustrados de forma gráfica, conforme abaixo:



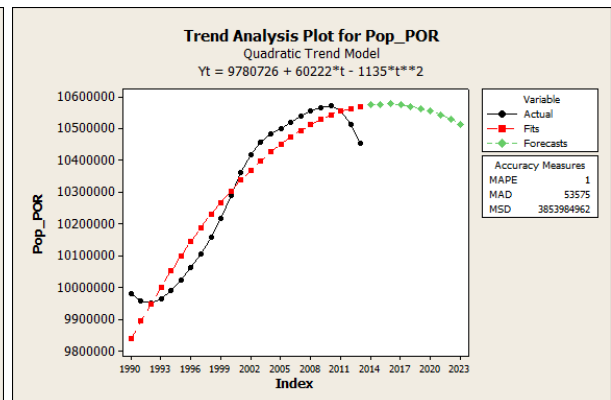
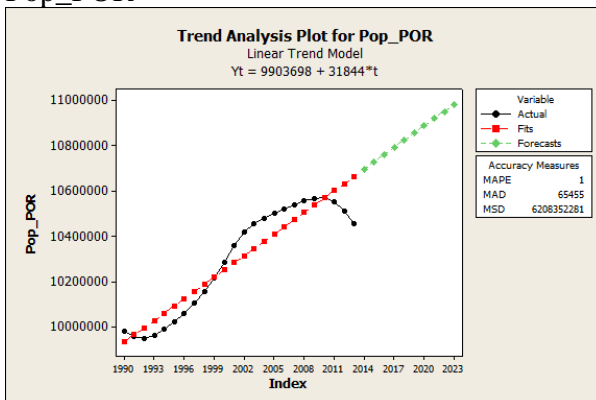
No gráfico acima, trata-se de uma previsão efetuada somente com base nos valores históricos da própria variável. Certamente esta metodologia apresenta deficiências ao assumir que o comportamento passado se repetirá. Do mesmo modo, talvez a utilização de regressões múltiplas poderiam ser mais eficientes neste trabalho de previsão.

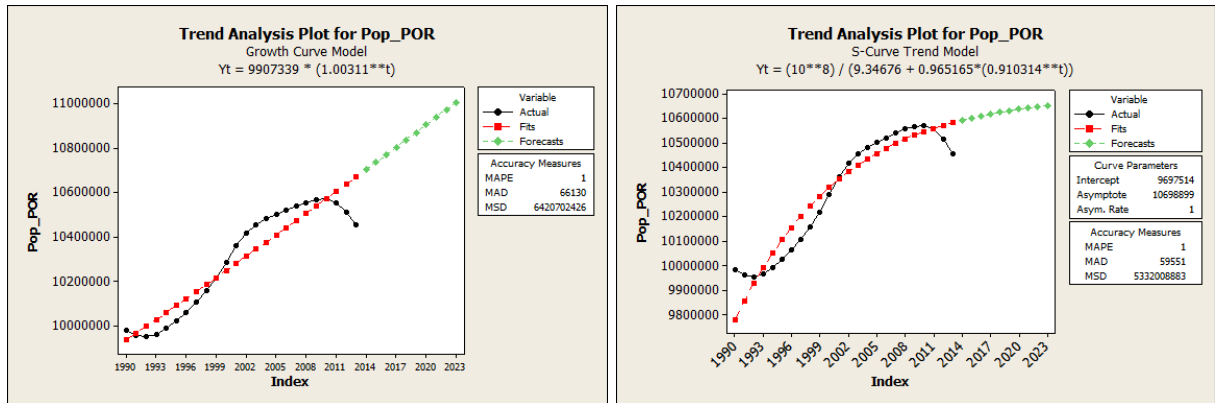
4.3.11 Variável: Pop_POR – População Total em Portugal

4.3.11.1. A análise de tendência

Segue abaixo gráficos incluindo a função, a linha de tendência que representa cada função e os erros mencionados acima.

Pop_POR





Para podermos visualizar qual a melhor função para representar a série de dados acima, podemos analisar a tabela abaixo, que compara os três tipos de erros para cada uma das quatro funções:

Estadística	Linear	Quadrática	Exponencial	S-Curve
MAPE	1	1	1	1
MAD	65455	53575	66130	59551
MSD	6208352281	3853984962	6420702426	5332008883

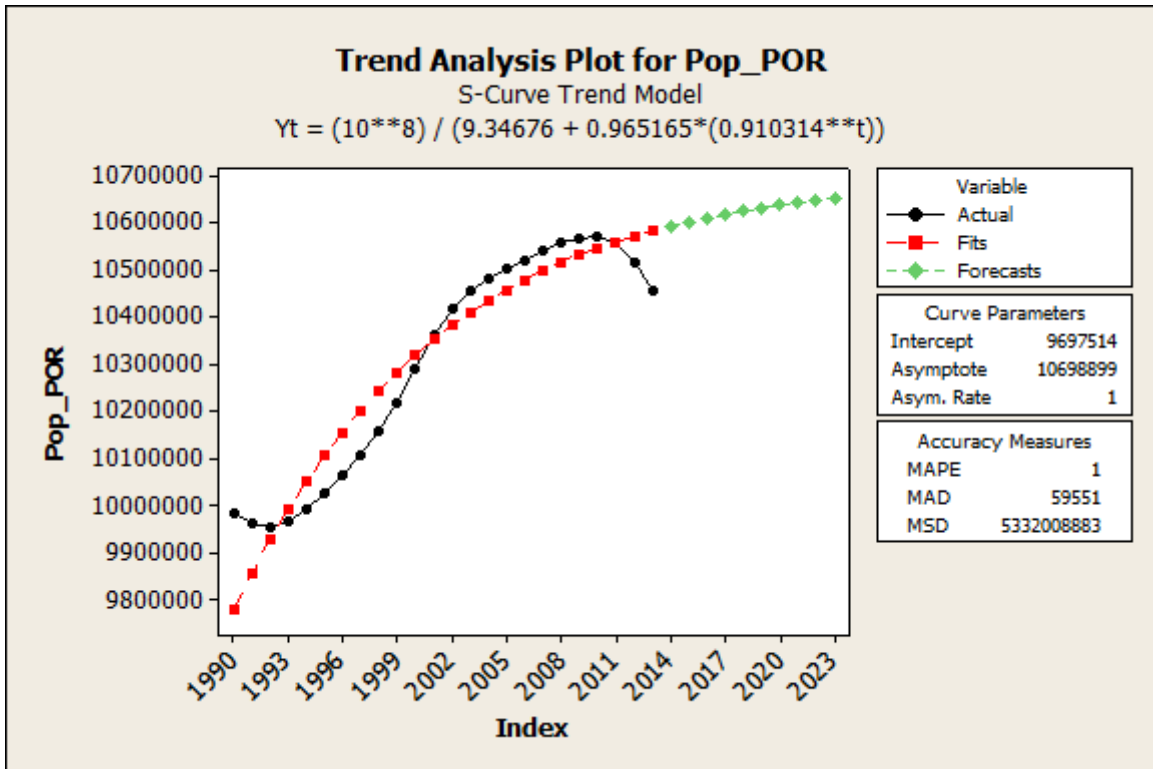
Percebemos que o modelo **Quadrática** é o que melhor se adapta à nossa série de dados. Assim, esta será a função utilizada para as projeções (*forecasts*).

4.3.11.2. As previsões

Utilizando a equação da função **Quadrática** apresentada no gráfico acima, temos os seguintes valores previstos (*forecasts*) para os próximos 10 anos em relação população total, em Portugal.

Year	Forecast
2014	10594475
2015	10603757
2016	10612221
2017	10619937
2018	10626971
2019	10633383
2020	10639226
2021	10644551
2022	10649402
2023	10653823

Os valores acima podem ser melhor ilustrados de forma gráfica, conforme abaixo:



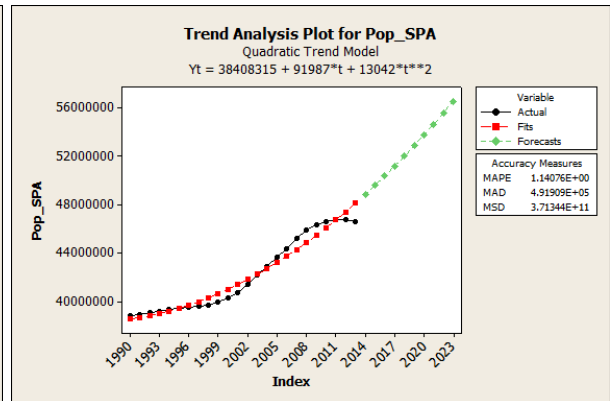
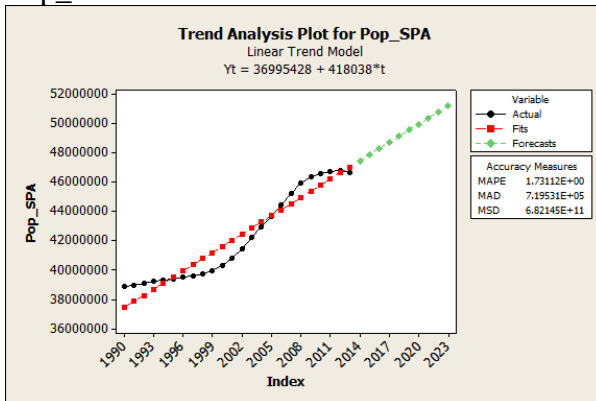
Ressaltamos que trata-se de uma previsão efetuada somente com base nos valores históricos da própria variável. Certamente esta metodologia apresenta deficiências ao assumir que o comportamento passado se repetirá. Do mesmo modo, talvez a utilização de regressões múltiplas poderiam ser mais eficientes neste trabalho de previsão.

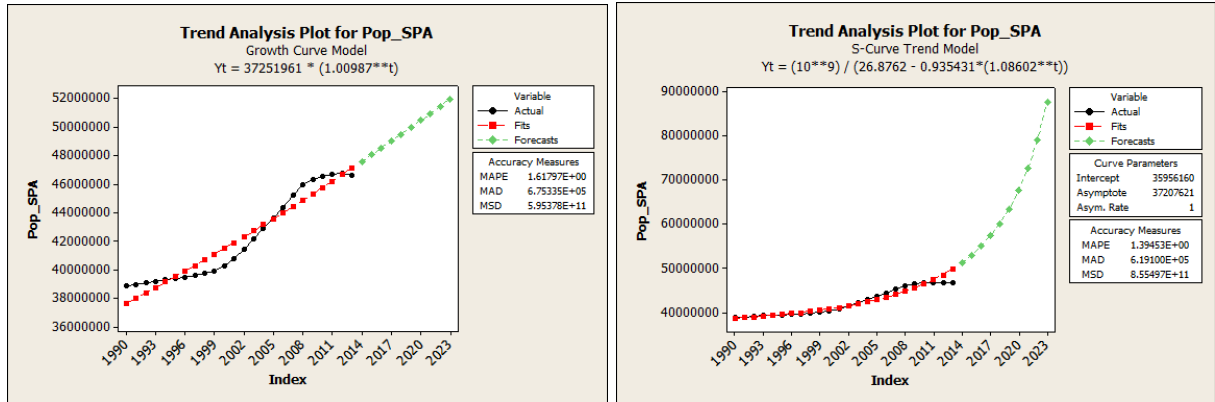
4.3.12 Variável: Pop_SPA – População Total na Espanha

4.3.12.1. A análise de tendência

Segue abaixo gráficos incluindo a função, a linha de tendência que representa cada função e os erros mencionados acima.

Pop_SPA





Para podermos visualizar qual a melhor função para representar a série de dados acima, podemos analisar a tabela abaixo, que compara os três tipos de erros para cada uma das quatro funções:

Estadística	Linear	Quadrática	Exponencial	S-Curve
MAPE	1.73112E+00	1.14076E+00	1.61797E+00	1.39453E+00
MAD	7.19531E+05	4.91909E+05	6.75335E+05	6.19100E+05
MSD	6.82145E+11	3.71344E+11	5.95378E+11	8.55497E+11

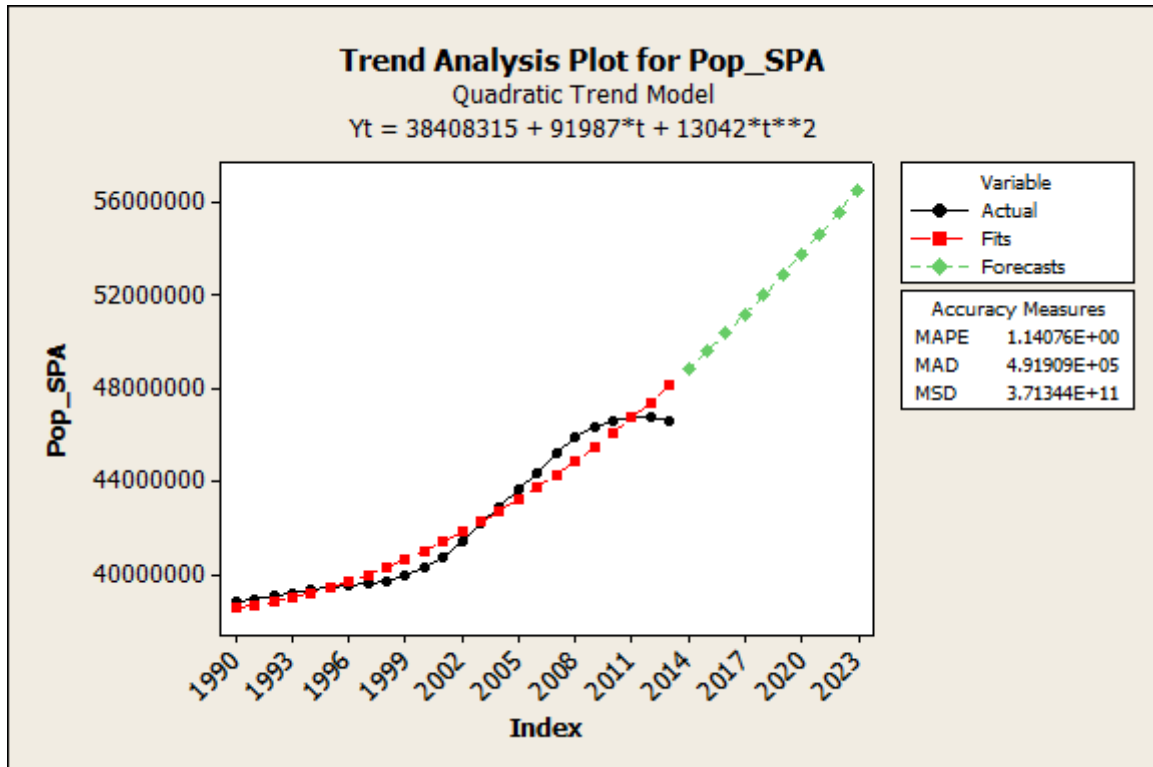
Notamos que o modelo **Quadrática** é o que melhor se adapta à nossa série de dados. Assim, esta será a função utilizada para as projeções (*forecasts*).

4.3.12.2. As previsões

Utilizando a equação da função **Quadrática** apresentada no gráfico acima, temos os seguintes valores previstos (*forecasts*) para os próximos 10 anos em relação a população total na Espanha.

Year	Forecast
2014	48859268
2015	49616399
2016	50399615
2017	51208914
2018	52044297
2019	52905764
2020	53793316
2021	54706951
2022	55646671
2023	56612474

Os valores acima podem ser melhor ilustrados de forma gráfica, conforme abaixo:



No gráfico acima, trata-se de uma previsão efetuada somente com base nos valores históricos da própria variável. Certamente esta metodologia apresenta deficiências ao assumir que o comportamento passado se repetirá. Do mesmo modo, talvez a utilização de regressões múltiplas poderiam ser mais eficientes neste trabalho de previsão.

5. COMPARAÇÕES

Nesta sessão será efetuada análise comparativa de médias e intervalos de confiança de dados dimensionadores. O principal propósito é comparar como os dados se comportarão quando agrupados em 3 principais regiões, os países Ibero-Americanos (AIBER), países de economias avançadas (AVECO) e os demais países constantes na base de dados deste estudo.

5.1. Categorização das variáveis – Siglas e Nomenclaturas

O objetivo é comparar a região Ibero-Americana (AIBER) com as demais regiões do mundo, para cada variável. Assim, cada variável será categorizada por região e receberá as siglas e nomenclaturas conforme segue:

Quadro - Siglas, Nomenclaturas e Quantidades de Países por Região (8)

Sigla da Região	Nomenclatura da Região	# Países
AIBER	Ibero-Americanos	21
AVECO	Economias Avançadas	27
CARLA	América Latina e o Caribe	3
CWEAS	Comunidade dos Estados Independentes	11
ESEAS	Emergentes e em Desenvolvimento da Ásia	12
EMDEU	Emergentes e em Desenvolvimento da Europa	12
NSWUA	Oriente Médio, Norte da África e Paquistão	14
SUSAF	África Subsariana	32

Total de países neste estudo: 132

Para que houvesse destaque visual e facilidade em identificar os agrupamentos das regiões, optou-se em colorir as células acima da seguinte forma:

- em azul, a região AIBER, contendo 21 países Ibero-Americanos;
- em verde, a região AVECO, que denota os 27 países considerados economias avançadas
- em vermelho, figuram 84 países das demais 6 regiões, que serão categorizados como OTHERS nas análises abaixo.

Obs: Conforme orientação recebida em sala, o país Cuba estava como CARLA (Caribbean and Latin America) e passou a figurar como AIBER, para fins deste estudo.

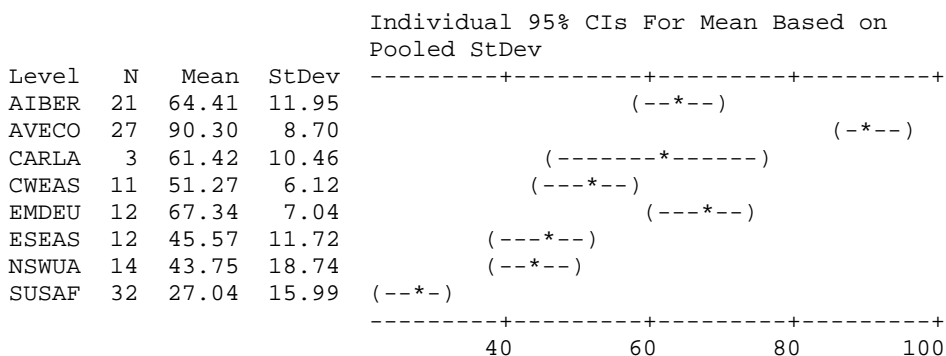
5.3. Comparação das regiões

Para esta análise comparativa será utilizada o modelo ANOVA. Esta análise de variância testa a hipótese de que as médias de duas ou mais populações são iguais.

One-way ANOVA: Social Progress Index_n versus Class 6R

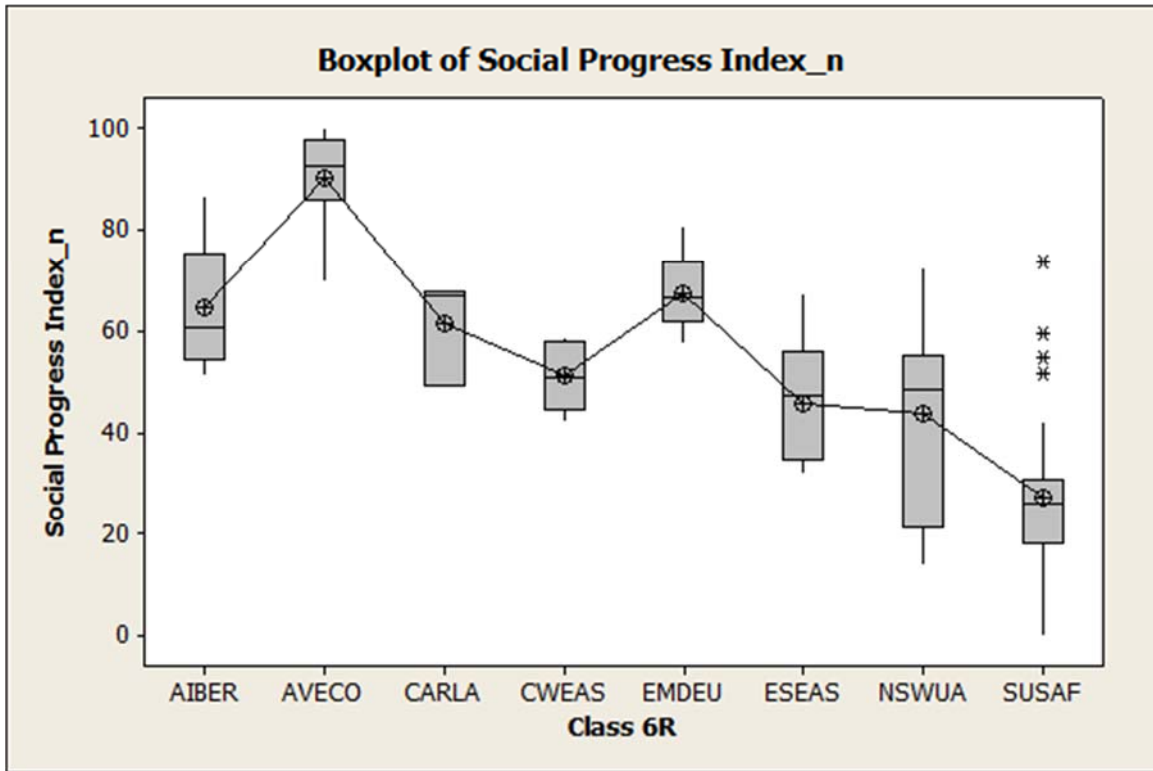
Source	DF	SS	MS	F	P
Class 6R	7	65372	9339	58.02	0.000
Error	124	19961	161		
Total	131	85333			

S = 12.69 R-Sq = 76.61% R-Sq(adj) = 75.29%



Pooled StDev = 12.69

Boxplot of Social Progress Index_n



One-way ANOVA: Social Progress Index_n versus Class 3R

Source	DF	SS	MS	F	P
Class 3R	2	48262	24131	83.97	0.000
Error	129	37071	287		
Total	131	85333			

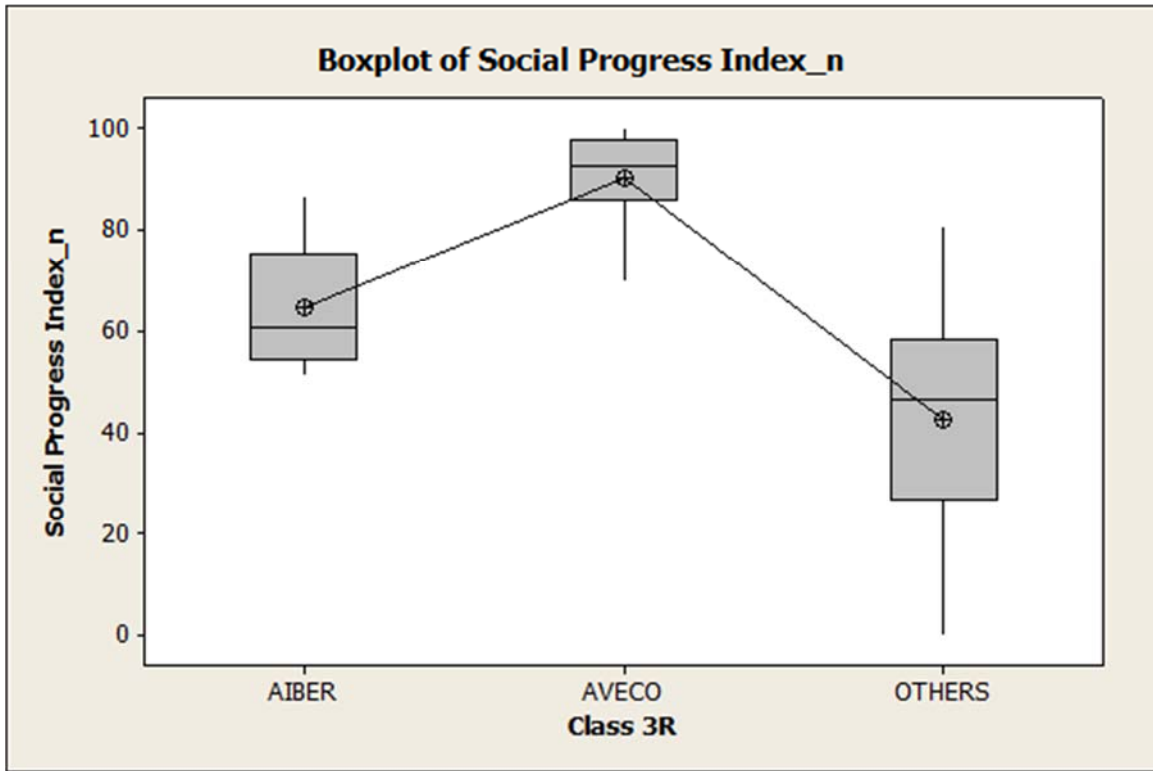
S = 16.95 R-Sq = 56.56% R-Sq(adj) = 55.88%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	CI Lower	CI Upper
AIBER	21	64.41	11.95	45.00	83.82
AVECO	27	90.30	8.70	72.90	107.70
OTHERS	84	42.63	19.71	12.21	73.05

Pooled StDev = 16.95

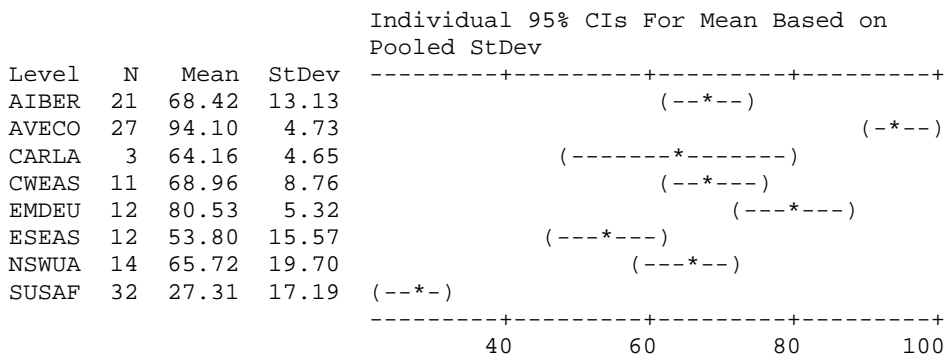
Boxplot of Social Progress Index_n



One-way ANOVA: Basic Human Needs_n versus Class 6R

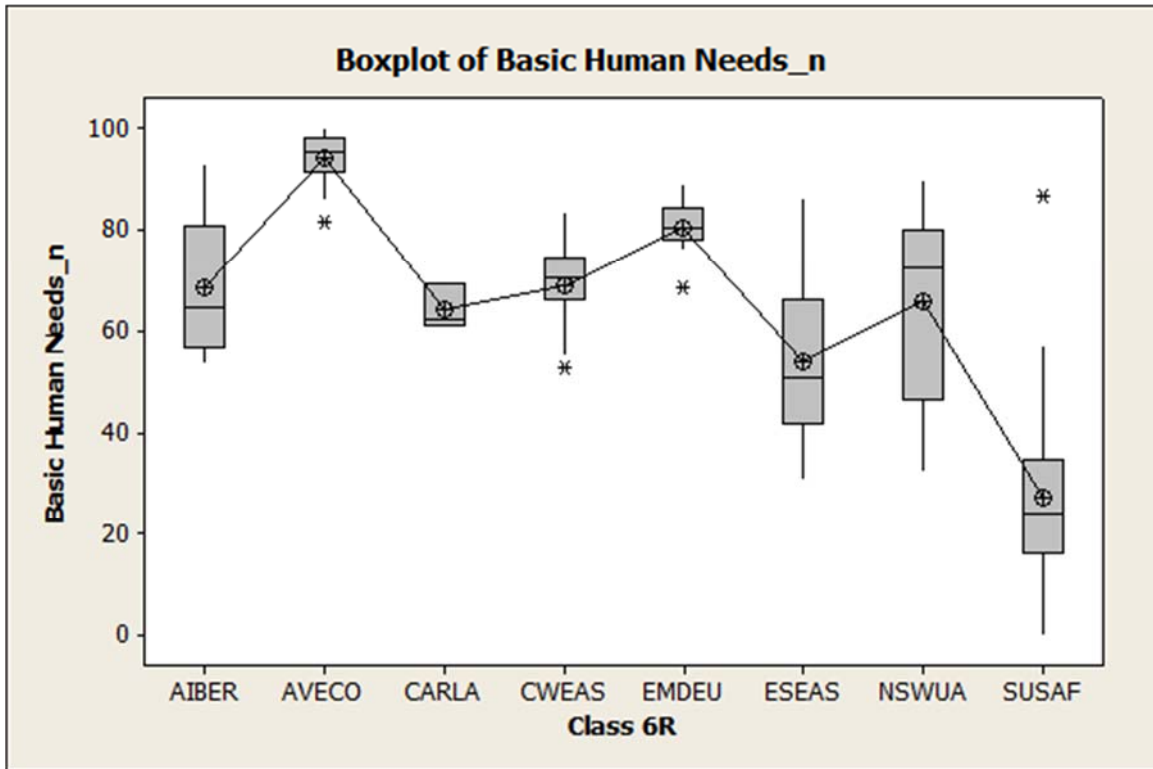
Source	DF	SS	MS	F	P
Class 6R	7	72694	10385	58.47	0.000
Error	124	22025	178		
Total	131	94719			

S = 13.33 R-Sq = 76.75% R-Sq(adj) = 75.43%



Pooled StDev = 13.33

Boxplot of Basic Human Needs_n



One-way ANOVA: Basic Human Needs_n versus Class 3R

Source	DF	SS	MS	F	P
Class 3R	2	37131	18566	41.59	0.000
Error	129	57588	446		
Total	131	94719			

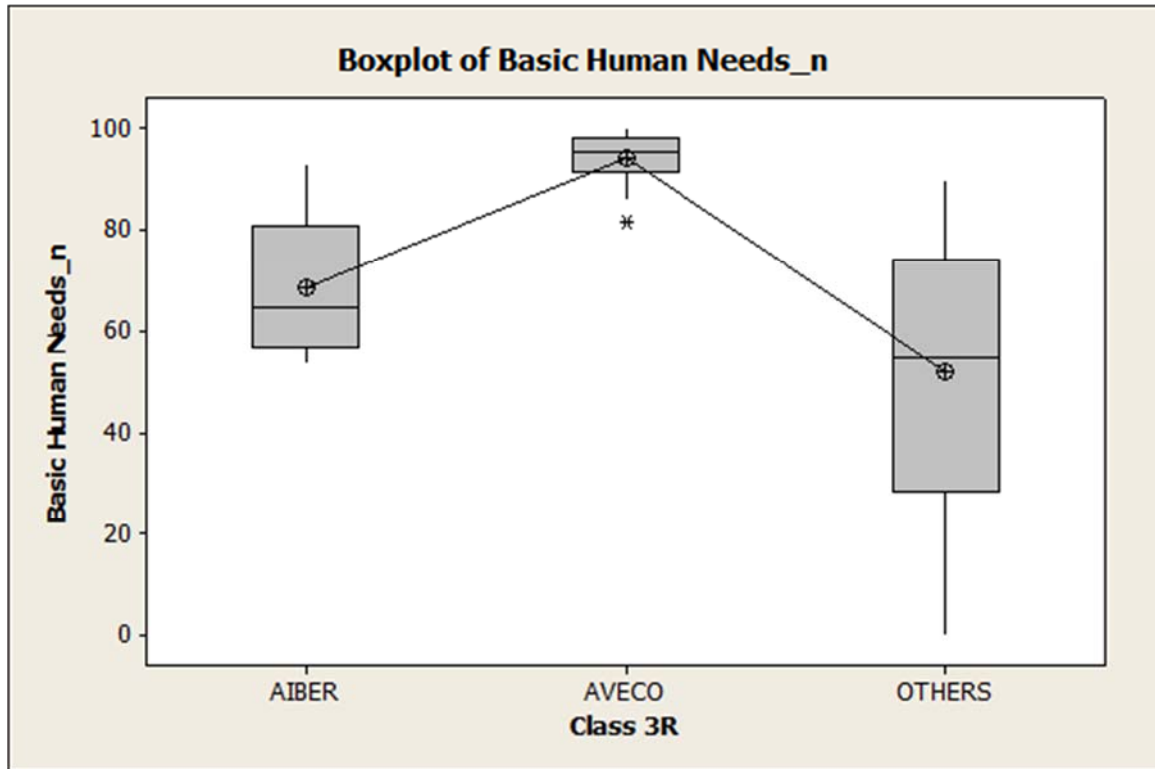
S = 21.13 R-Sq = 39.20% R-Sq(adj) = 38.26%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	CI Lower	CI Upper
AIBER	21	68.42	13.13	52.00	84.84
AVECO	27	94.10	4.73	84.64	103.56
OTHERS	84	51.87	25.40	0.07	103.67

Pooled StDev = 21.13

Boxplot of Basic Human Needs_n

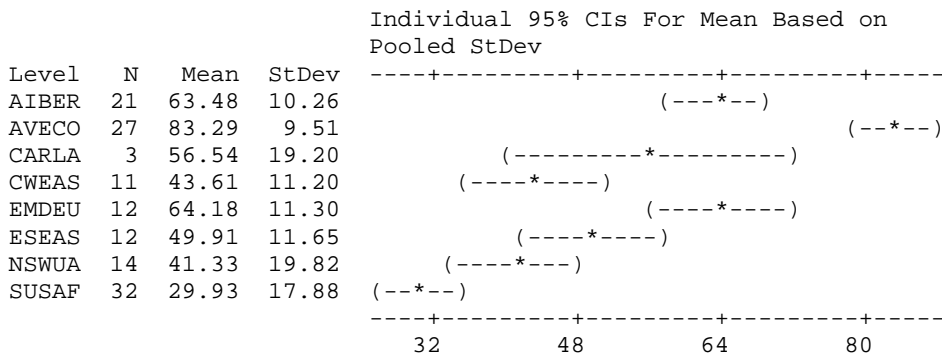


Obs: O país Latvia (Letônia) figura como *outlier* no agrupamento de países da região AVECO, para esta variável.

One-way ANOVA: Foundations of Wellbeing_n versus Class 6R

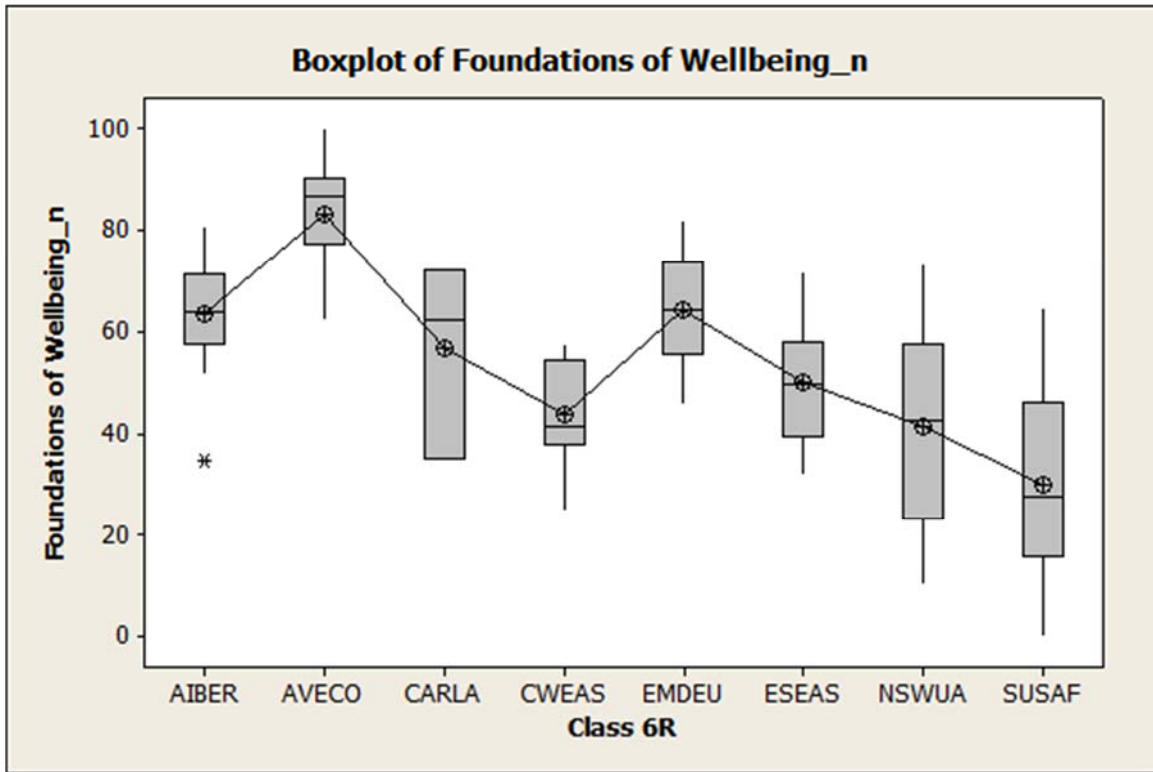
Source	DF	SS	MS	F	P
Class 6R	7	48501	6929	35.27	0.000
Error	124	24358	196		
Total	131	72859			

S = 14.02 R-Sq = 66.57% R-Sq(adj) = 64.68%



Pooled StDev = 14.02

Boxplot of Foundations of Wellbeing_n

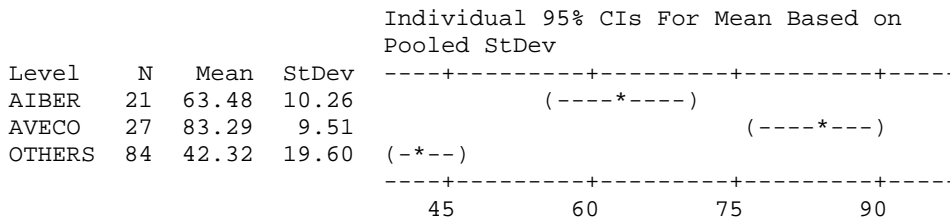


Obs: Os 3 países que compõem a região CARLA, são: Guiana, Jamaica e Trindade e Tobago

One-way ANOVA: Foundations of Wellbeing_n versus Class 3R

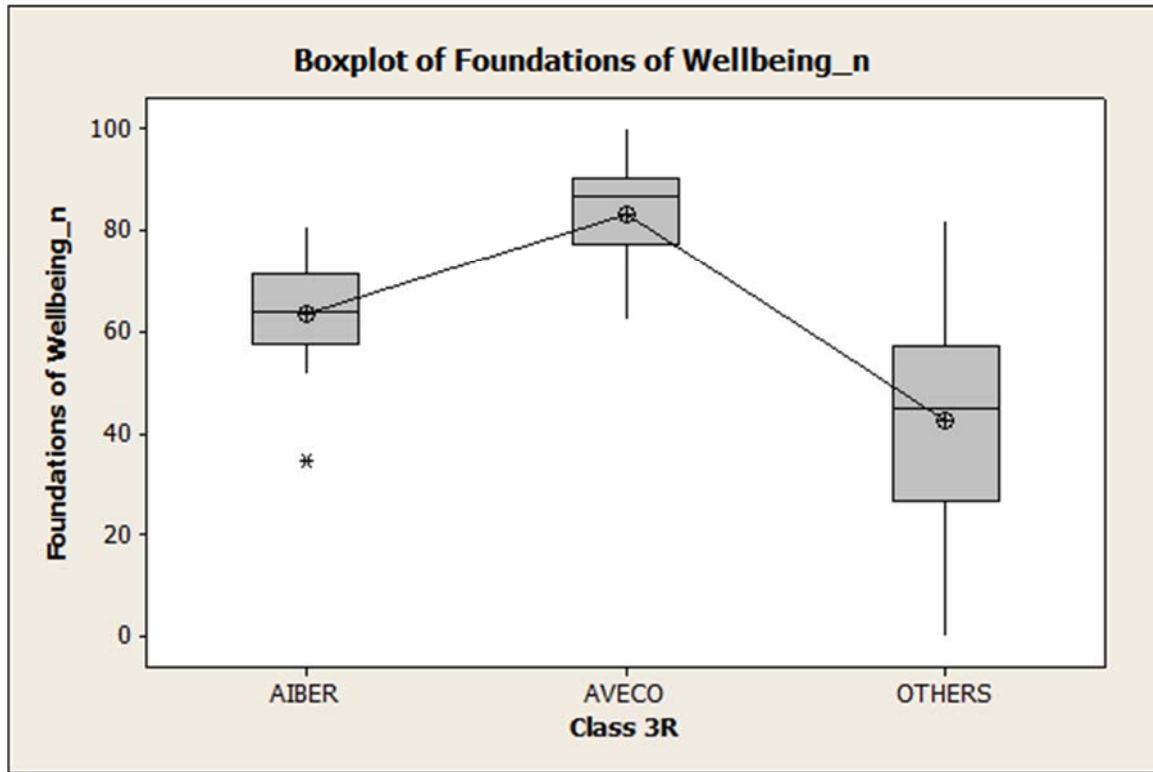
Source	DF	SS	MS	F	P
Class 3R	2	36523	18262	64.83	0.000
Error	129	36336	282		
Total	131	72859			

S = 16.78 R-Sq = 50.13% R-Sq(adj) = 49.36%



Pooled StDev = 16.78

Boxplot of Foundations of Wellbeing_n

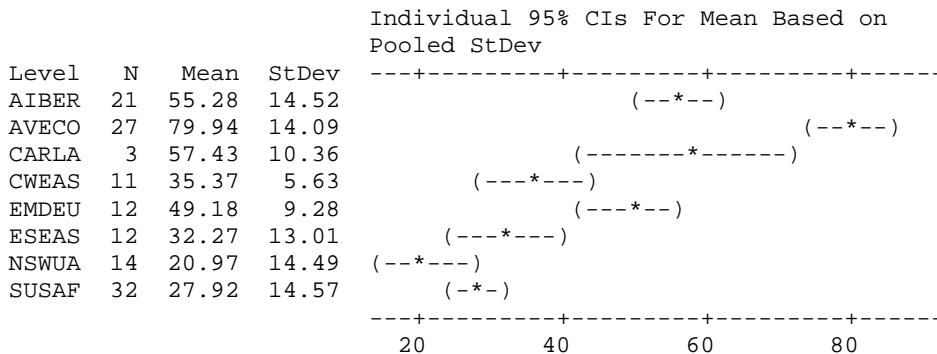


Obs: No gráfico acima, Cuba figura como *outlier* no agrupamento de países da região AIBER.

One-way ANOVA: Opportunity_n versus Class 6R

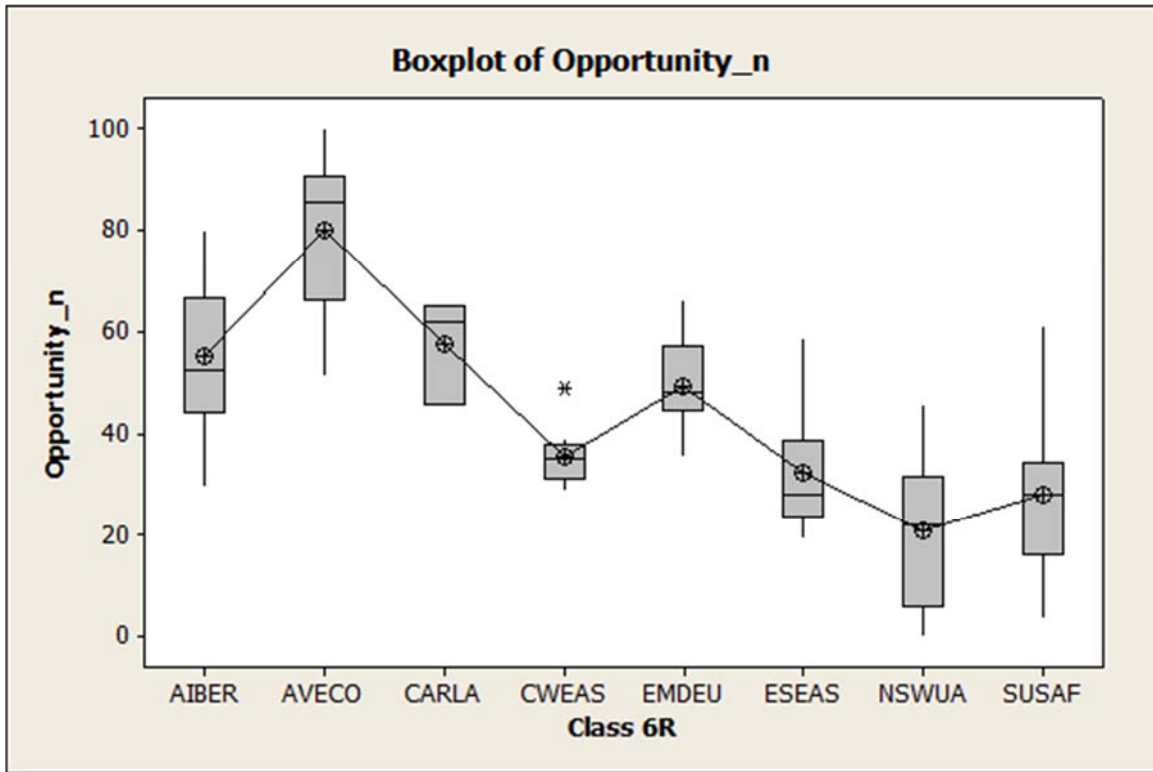
Source	DF	SS	MS	F	P
Class 6R	7	56157	8022	45.15	0.000
Error	124	22034	178		
Total	131	78191			

S = 13.33 R-Sq = 71.82% R-Sq(adj) = 70.23%



Pooled StDev = 13.33

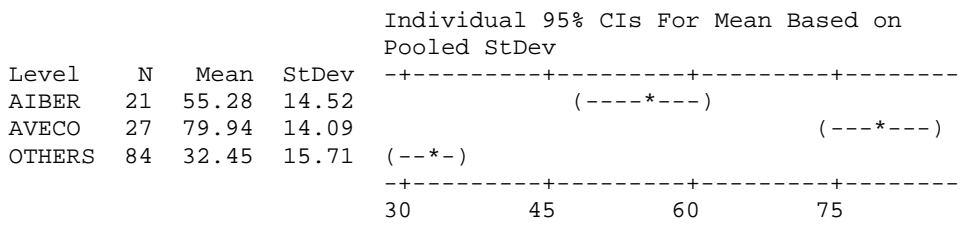
Boxplot of Opportunity_n



One-way ANOVA: Opportunity_n versus Class 3R

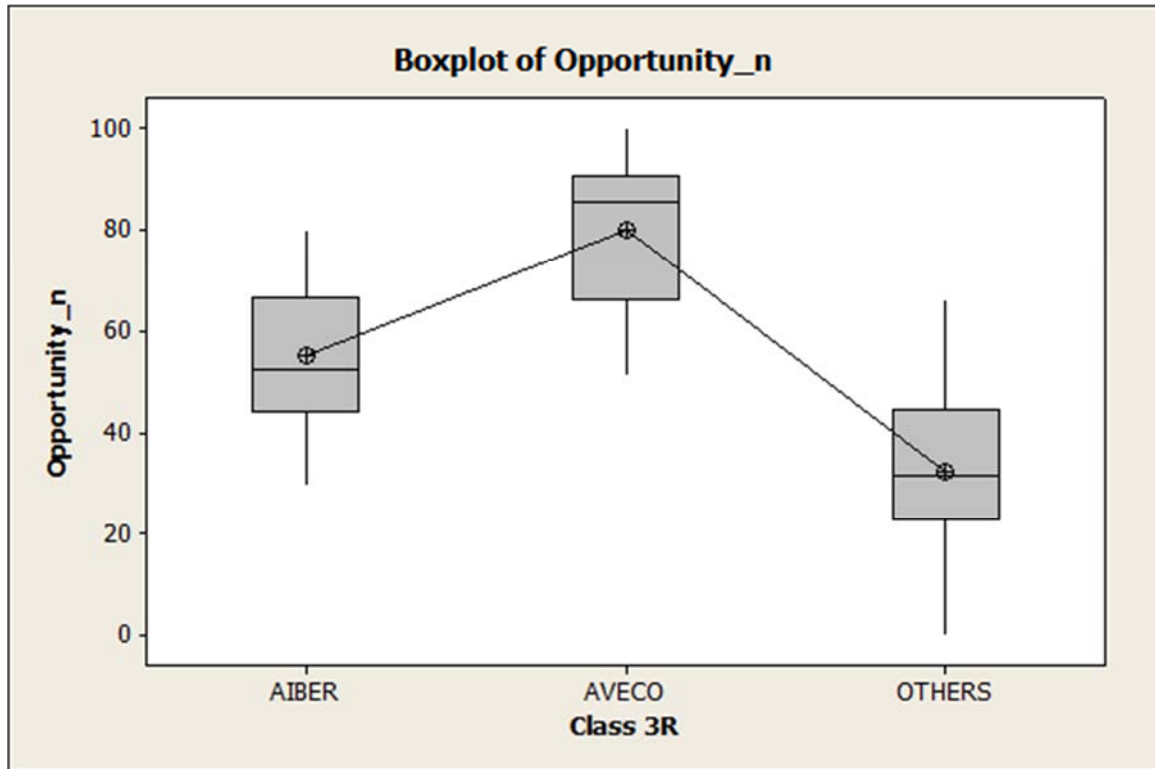
Source	DF	SS	MS	F	P
Class 3R	2	48330	24165	104.39	0.000
Error	129	29861	231		
Total	131	78191			

S = 15.21 R-Sq = 61.81% R-Sq(adj) = 61.22%



Pooled StDev = 15.21

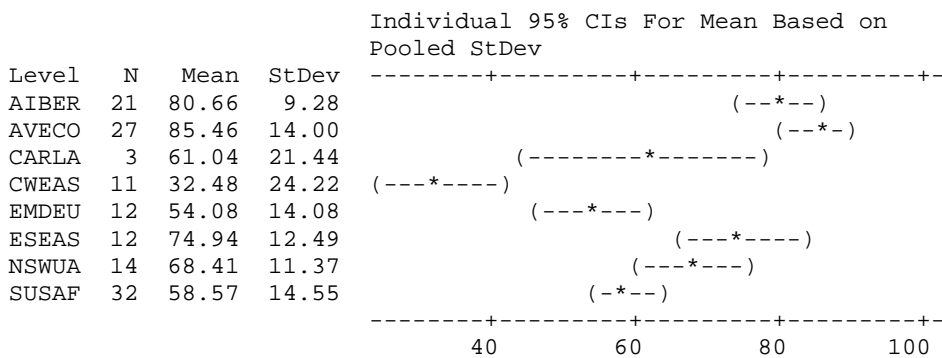
Boxplot of Opportunity_n



One-way ANOVA: Health and Wellness_n versus Class 6R

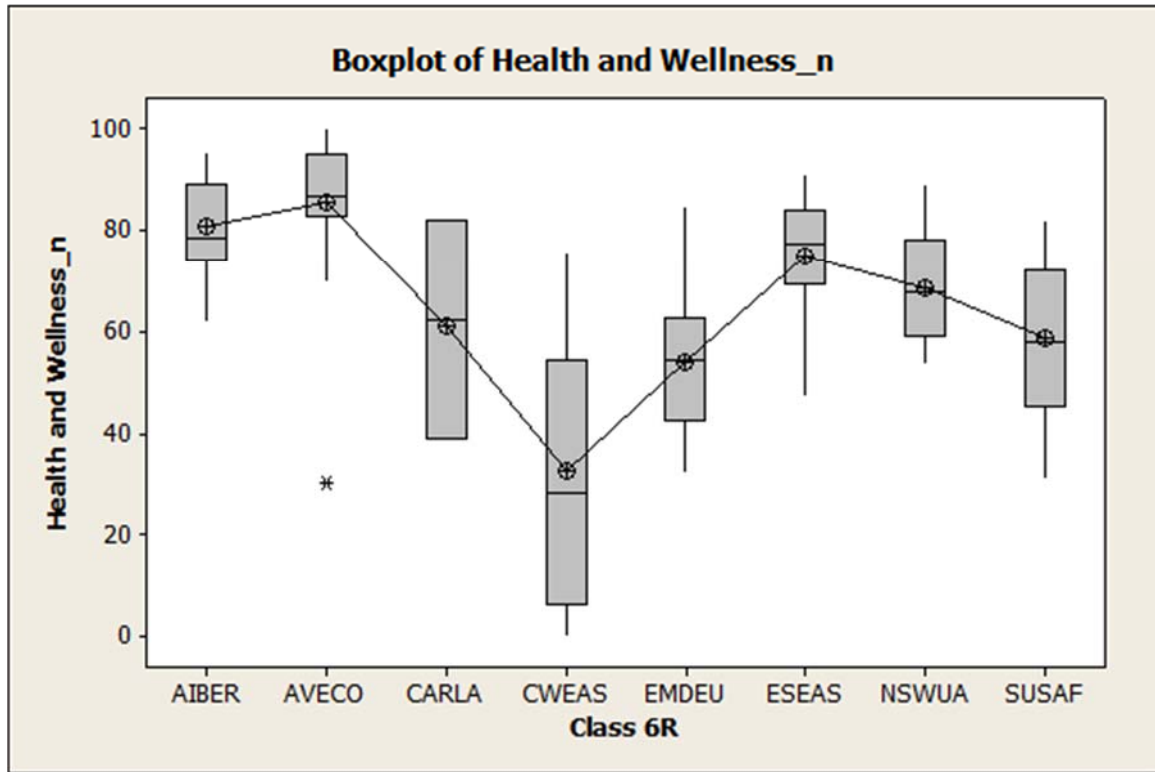
Source	DF	SS	MS	F	P
Class 6R	7	31347	4478	21.57	0.000
Error	124	25748	208		
Total	131	57095			

S = 14.41 R-Sq = 54.90% R-Sq(adj) = 52.36%



Pooled StDev = 14.41

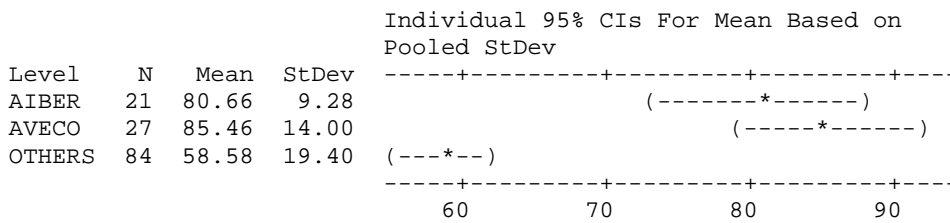
Boxplot of Health and Wellness_n



One-way ANOVA: Health and Wellness_n versus Class 3R

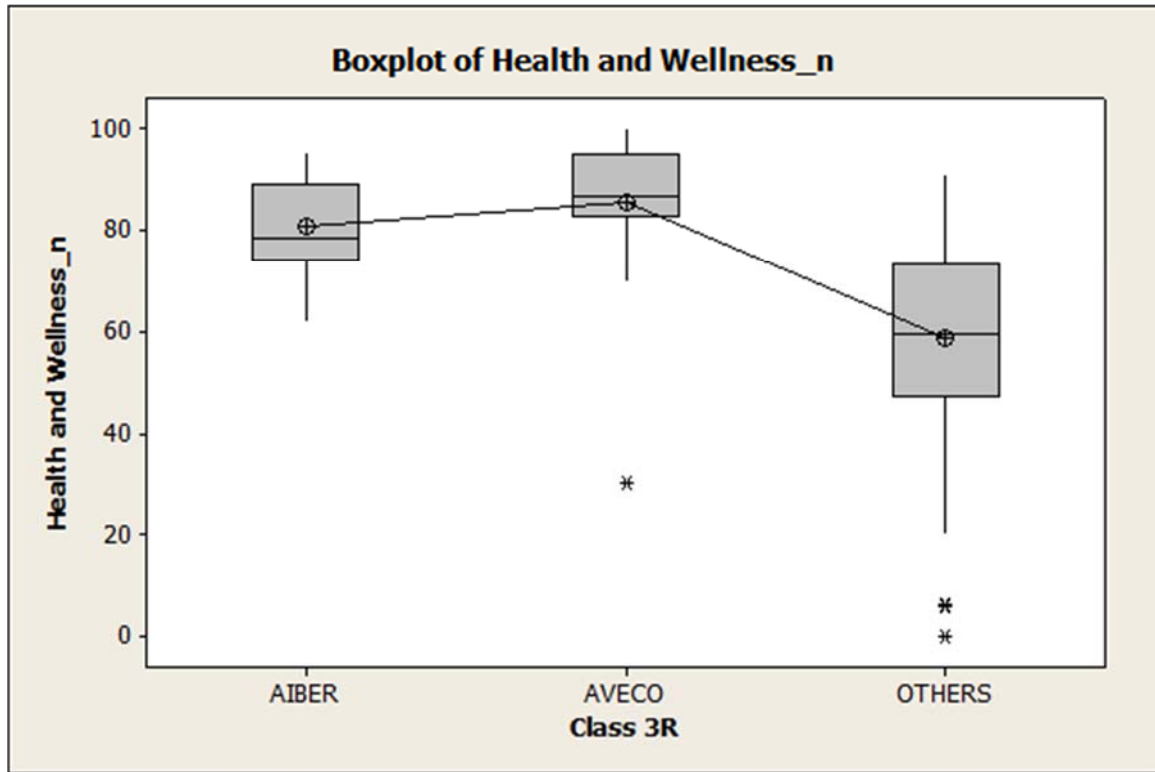
Source	DF	SS	MS	F	P
Class 3R	2	19025	9513	32.23	0.000
Error	129	38070	295		
Total	131	57095			

S = 17.18 R-Sq = 33.32% R-Sq(adj) = 32.29%



Pooled StDev = 17.18

Boxplot of Health and Wellness_n



Obs: A Letônia (Latvia), aparece como *outlier* nesta avaliação, denotando resultado bem abaixo dos de países do agrupamento AVECO. E no grupo OTHERS, os dois *outliers* são Rússia e Cazaquistão (Kazakhstan).

One-way ANOVA: Personal Freedom and Choice_n versus Class 6R

Source	DF	SS	MS	F	P
Class 6R	7	21207	3030	23.54	0.000
Error	124	15961	129		
Total	131	37168			

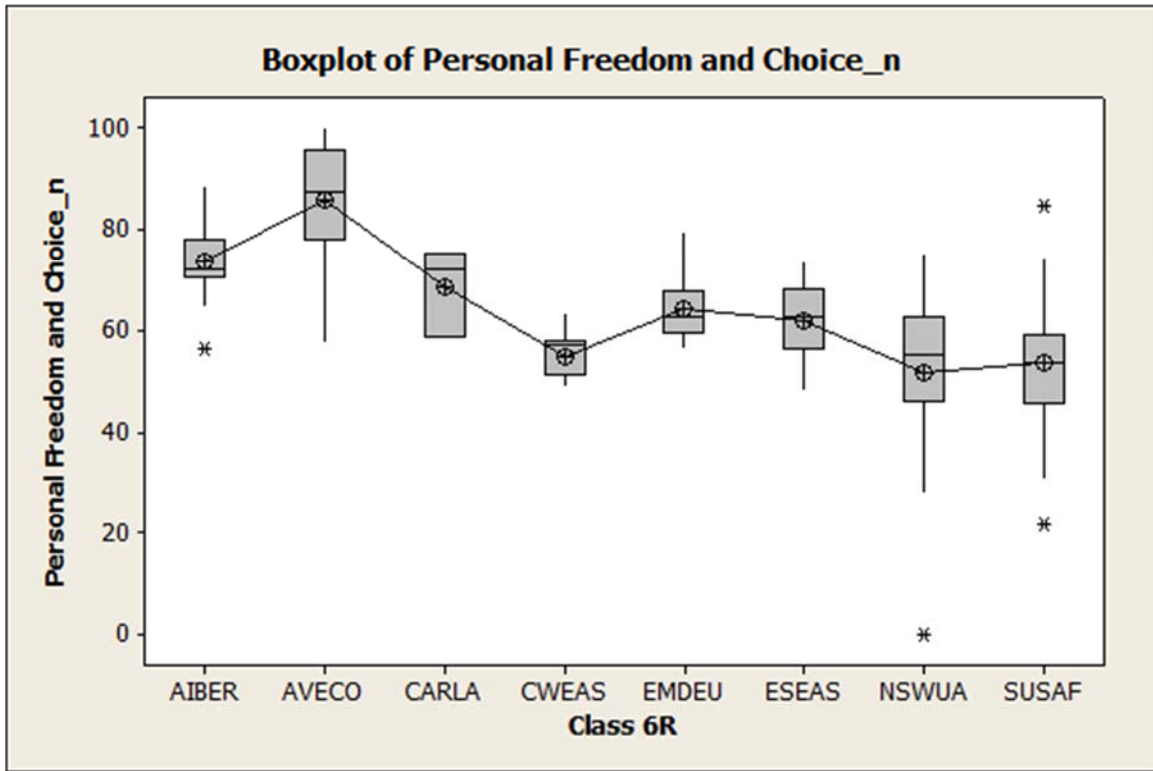
S = 11.35 R-Sq = 57.06% R-Sq(adj) = 54.63%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	CI Lower	CI Upper
AIBER	21	73.74	7.32	63.5	84.0
AVECO	27	85.82	11.92	71.0	100.6
CARLA	3	68.60	8.78	50.0	87.2
CWEAS	11	54.87	4.54	45.0	64.7
EMDEU	12	64.10	6.30	55.0	73.2
ESEAS	12	61.66	8.22	51.0	72.3
NSWUA	14	51.69	19.34	12.0	91.4
SUSAF	32	53.64	12.43	38.0	69.3

Pooled StDev = 11.35

Boxplot of Personal Freedom and Choice_n



One-way ANOVA: Personal Freedom and Choice_n versus Class 3R

Source	DF	SS	MS	F	P
Class 3R	2	19142	9571	68.49	0.000
Error	129	18027	140		
Total	131	37168			

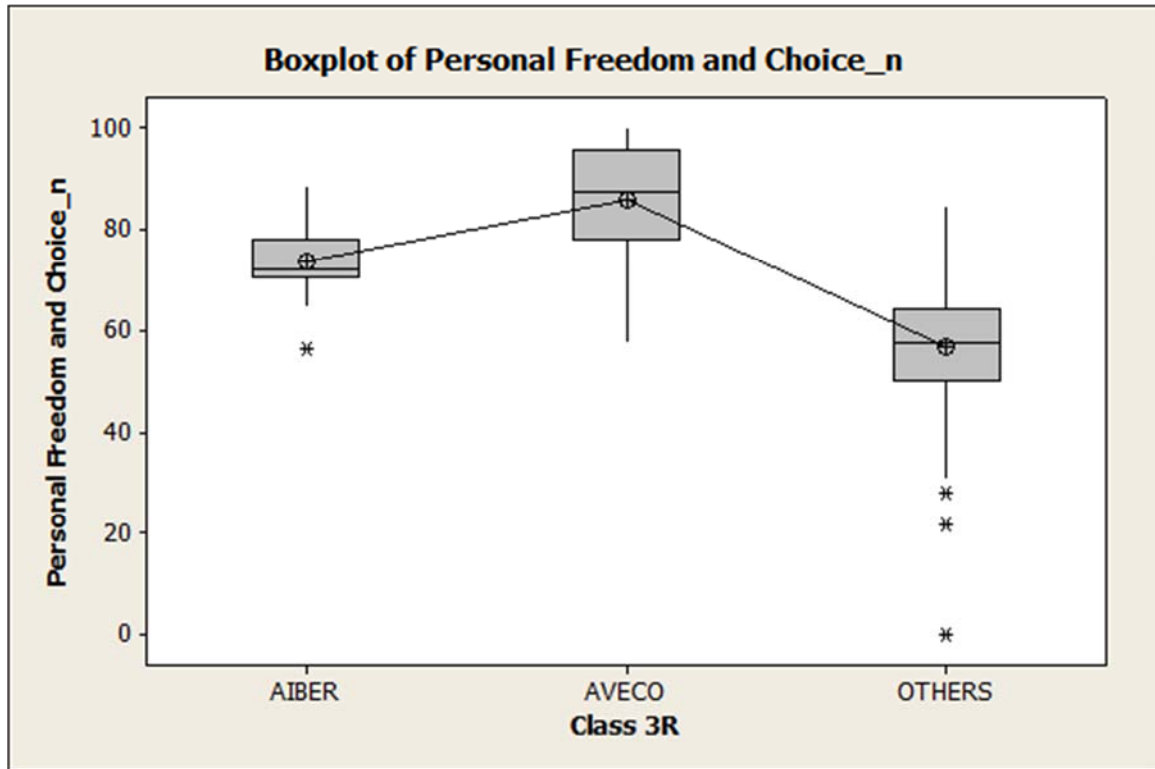
S = 11.82 R-Sq = 51.50% R-Sq(adj) = 50.75%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	CI Lower	CI Upper
AIBER	21	73.74	7.32	65.8	81.7
AVECO	27	85.82	11.92	72.8	98.8
OTHERS	84	56.65	12.64	43.9	69.4

Pooled StDev = 11.82

Boxplot of Personal Freedom and Choice_n



Obs: No agrupamento de países AIBER, Cuba aparece como *outlier* em relação aos demais. Os outliers do grupo OTHERS são respectivamente Paquistão, Sudão e Mauritânia.

One-way ANOVA: Greenhouse gas emissions_np versus Class 6R

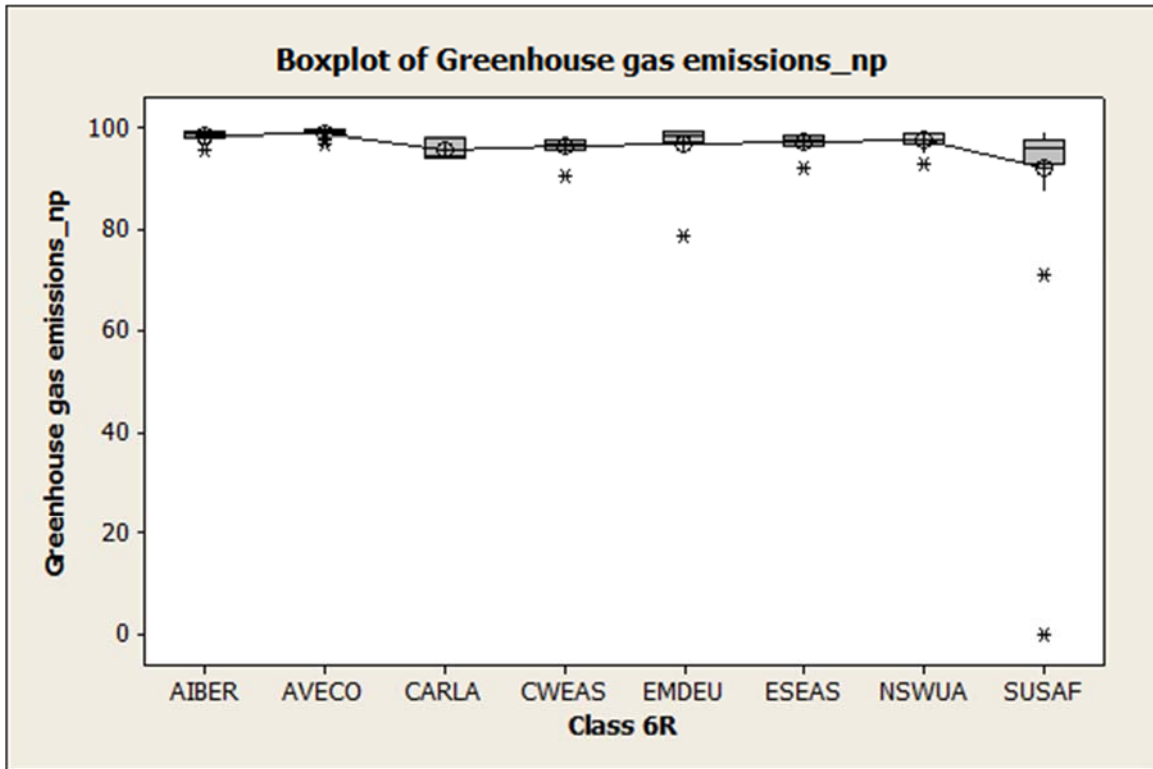
Source	DF	SS	MS	F	P
Class 6R	7	933.2	133.3	1.64	0.131
Error	124	10092.4	81.4		
Total	131	11025.6			

S = 9.022 R-Sq = 8.46% R-Sq(adj) = 3.30%

Level	N	Mean	StDev	Individual 95% CIs For Mean Based on Pooled StDev
AIBER	21	98.53	1.02	(-----*-----)
AVECO	27	99.09	0.62	(-----*-----)
CARLA	3	95.64	2.23	(-----*-----)
CWEAS	11	96.53	2.07	(-----*-----)
EMDEU	12	96.93	5.75	(-----*-----)
ESEAS	12	97.26	1.89	(-----*-----)
NSWUA	14	97.60	1.83	(-----*-----)
SUSAF	32	92.04	17.56	(-----*-----)

Pooled StDev = 9.02

Boxplot of Greenhouse gas emissions_np



One-way ANOVA: Greenhouse gas emissions_np versus Class 3R

Source	DF	SS	MS	F	P
Class 3R	2	426.8	213.4	2.60	0.078
Error	129	10598.8	82.2		
Total	131	11025.6			

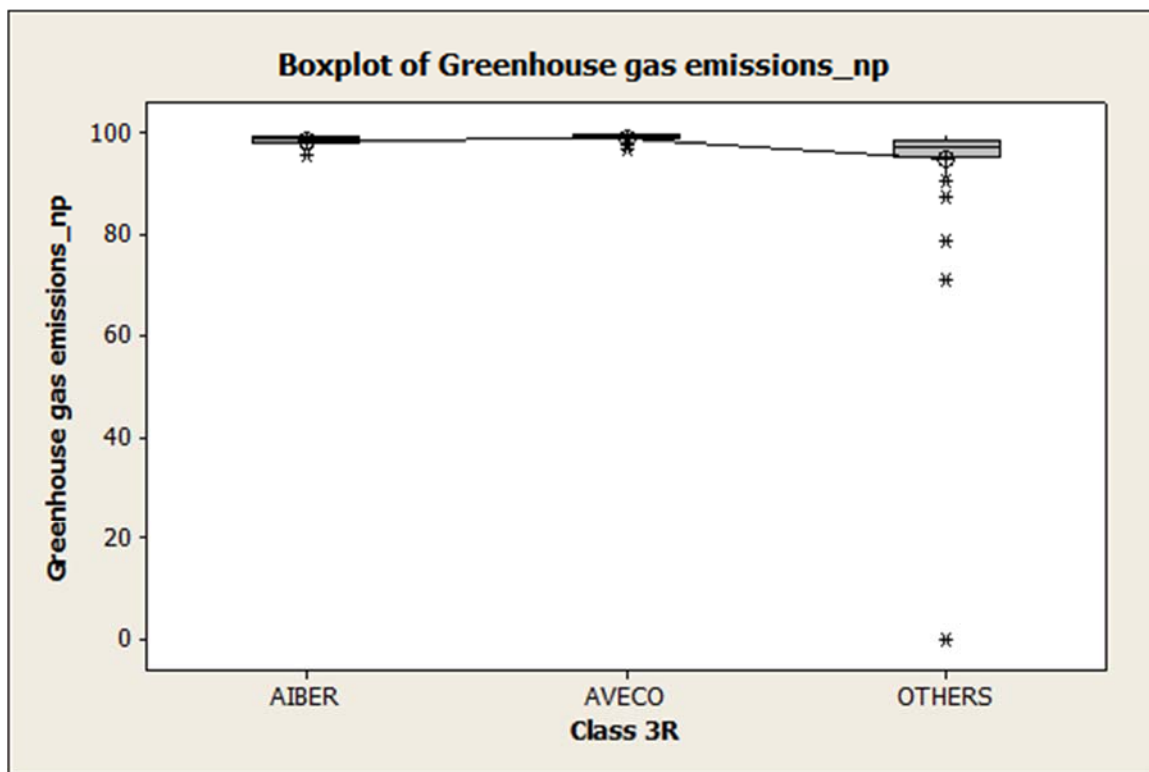
S = 9.064 R-Sq = 3.87% R-Sq(adj) = 2.38%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	CI Lower	CI Upper
AIBER	21	98.53	1.02	97.4	99.6
AVECO	27	99.09	0.62	98.4	99.7
OTHERS	84	95.13	11.28	83.8	106.5

Pooled StDev = 9.06

Boxplot of Greenhouse gas emissions_np

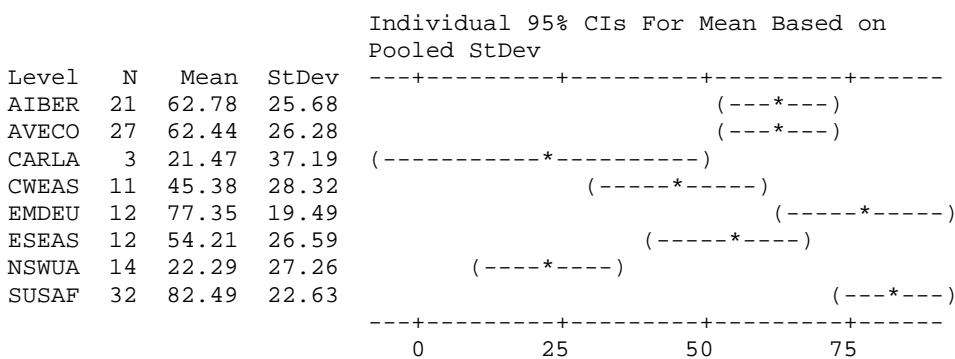


Obs:A República Centro-Africana (Central African Republic), no coração da África, aparece neste estudo como o maior emissor de gases de efeito estufa. Segundo o relatório da ONU, sobre os dados de 2010, a principal causa deste resultado estava ligado a Desmatamento e Exploração de madeira para comércio (89.46%), seguido Agricultura (5.26%), entre outros.

One-way ANOVA: Water withdrawals_np versus Class 6R

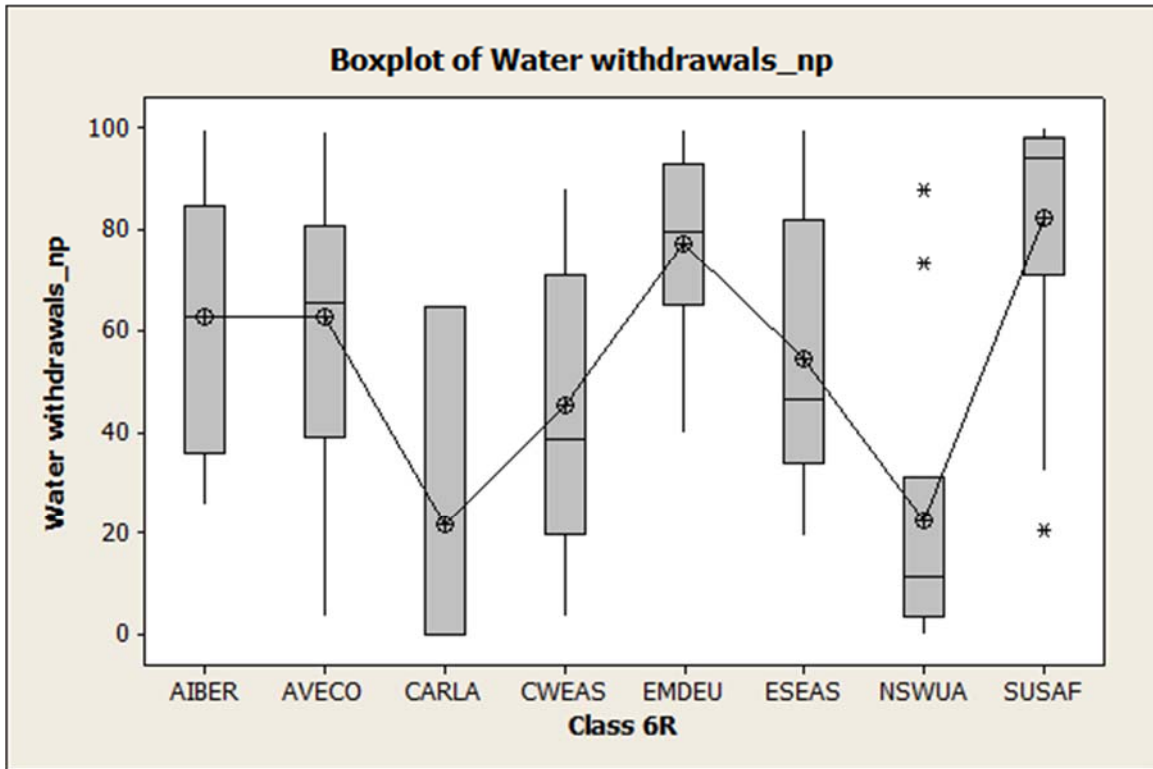
Source	DF	SS	MS	F	P
Class 6R	7	46991	6713	10.48	0.000
Error	124	79426	641		
Total	131	126417			

S = 25.31 R-Sq = 37.17% R-Sq(adj) = 33.62%



Pooled StDev = 25.31

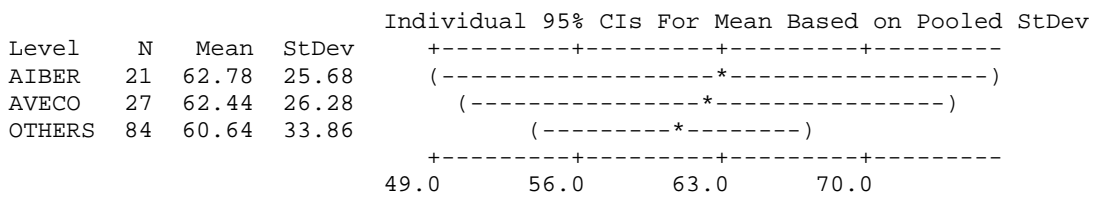
Boxplot of Water withdrawals_np



One-way ANOVA: Water withdrawals_np versus Class 3R

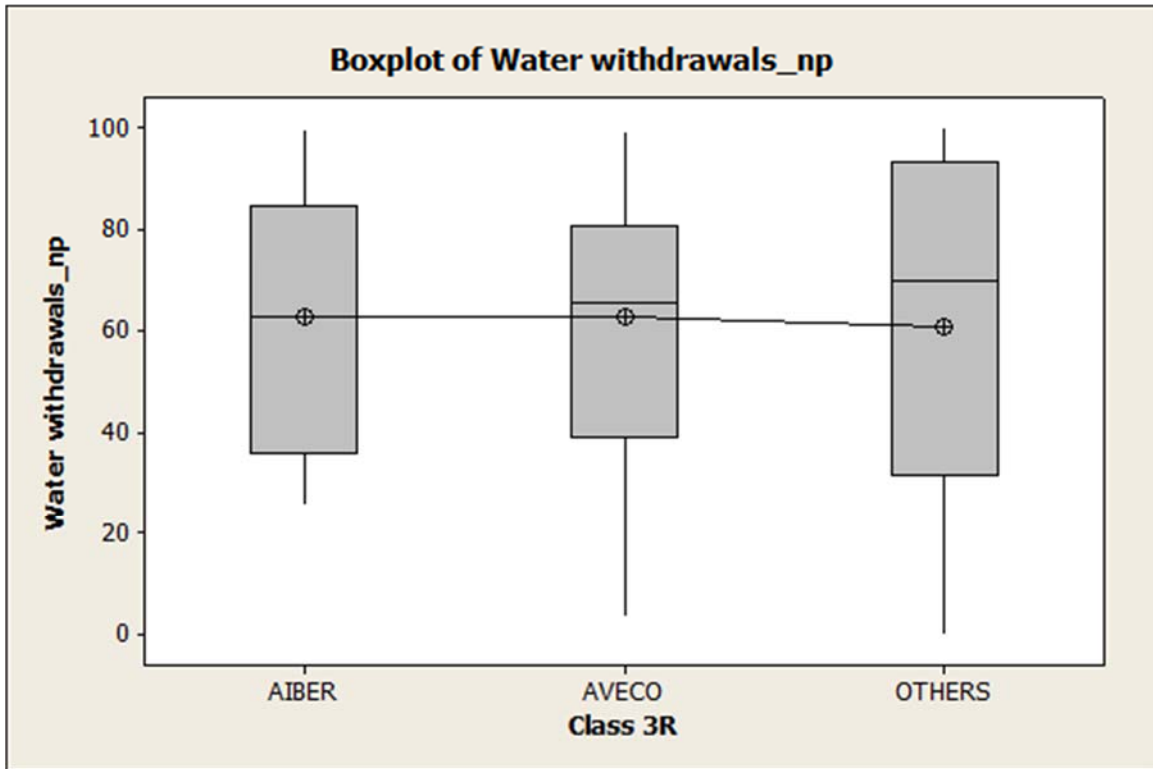
Source	DF	SS	MS	F	P
Class 3R	2	117	58	0.06	0.942
Error	129	126300	979		
Total	131	126417			

S = 31.29 R-Sq = 0.09% R-Sq(adj) = 0.00%



Pooled StDev = 31.29

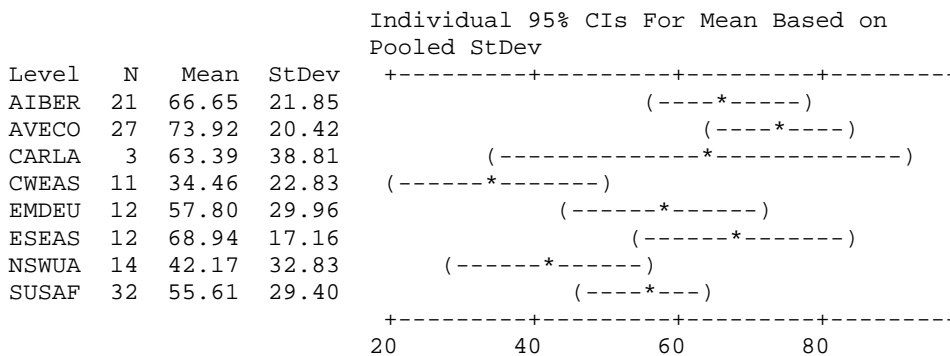
Boxplot of Water withdrawals_np



One-way ANOVA: Biodiversity and habitat_n versus Class 6R

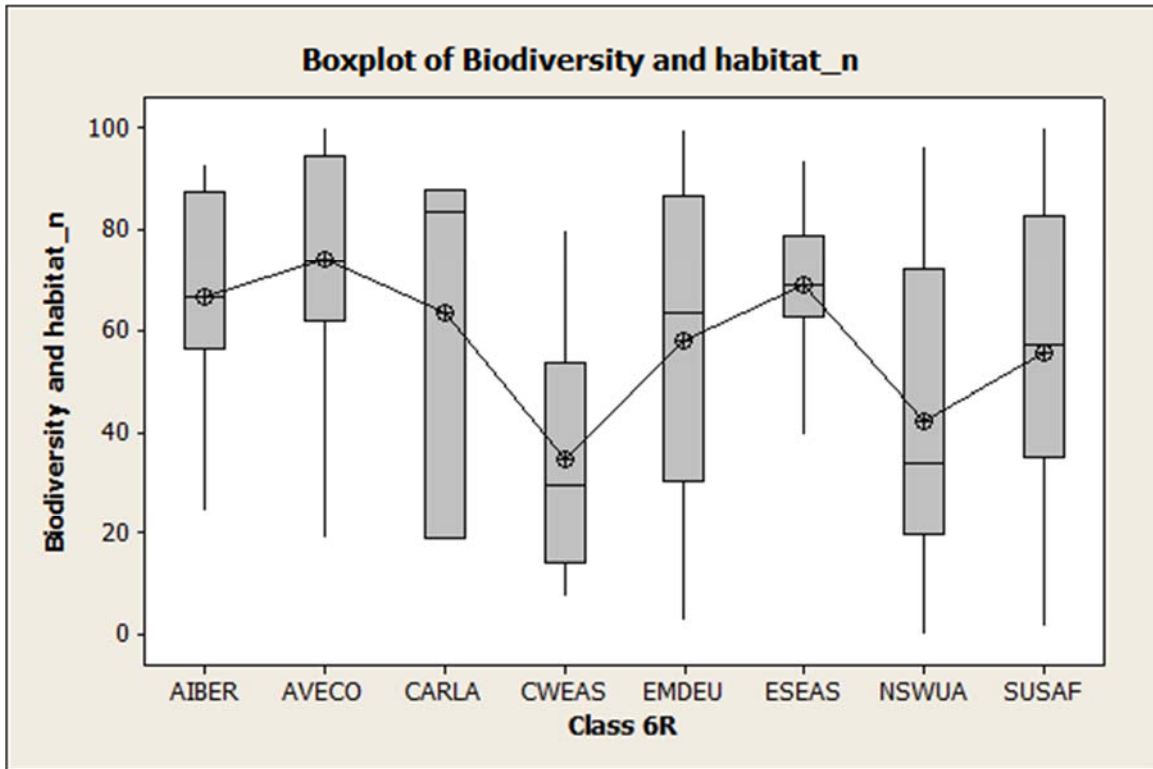
Source	DF	SS	MS	F	P
Class 6R	7	19429	2776	4.17	0.000
Error	124	82530	666		
Total	131	101958			

S = 25.80 R-Sq = 19.06% R-Sq(adj) = 14.49%



Pooled StDev = 25.80

Boxplot of Biodiversity and habitat_n



One-way ANOVA: Biodiversity and habitat_n versus Class 3R

Source	DF	SS	MS	F	P
Class 3R	2	10137	5069	7.12	0.001
Error	129	91821	712		
Total	131	101958			

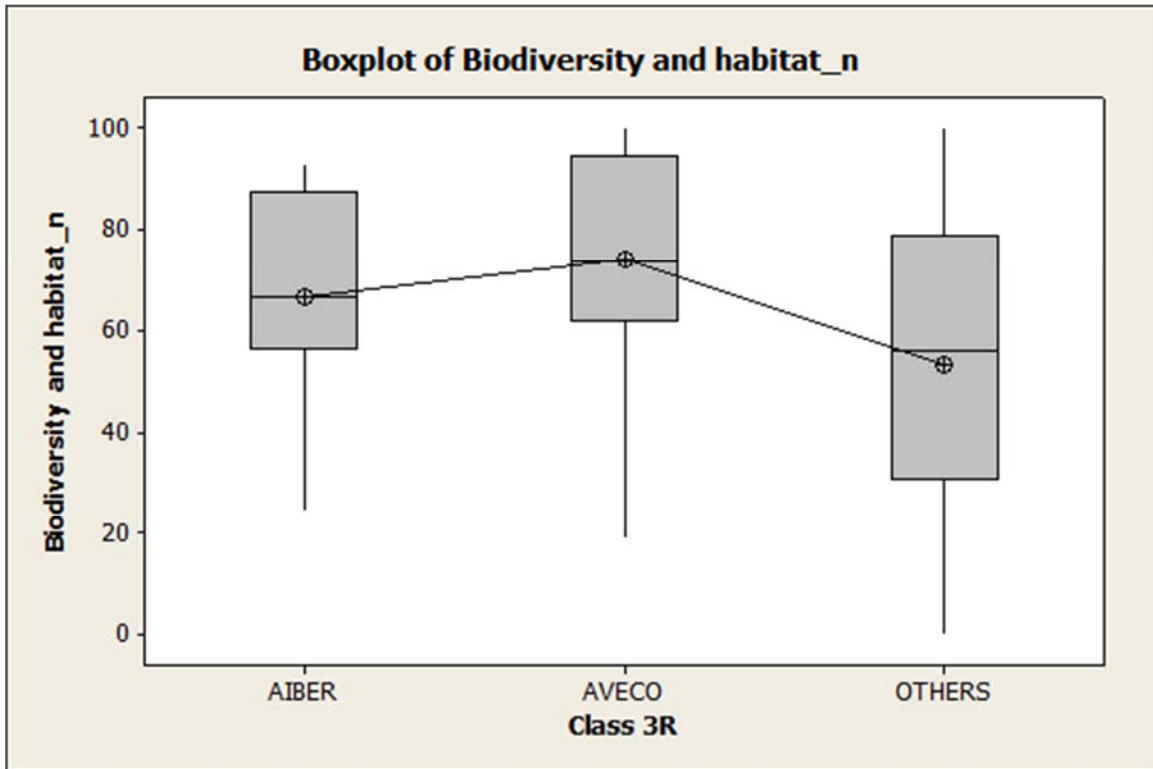
S = 26.68 R-Sq = 9.94% R-Sq(adj) = 8.55%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev
AIBER	21	66.65	21.85
AVECO	27	73.92	20.42
OTHERS	84	53.10	29.34

Pooled StDev = 26.68

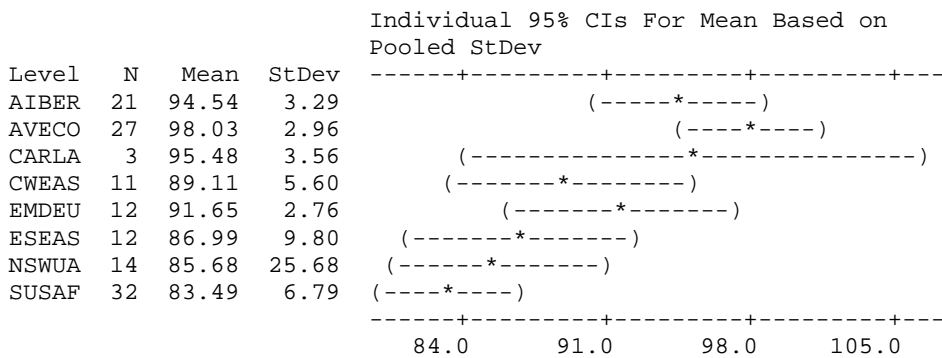
Boxplot of Biodiversity and habitat_n



One-way ANOVA: Modern slavery_np versus Class 6R

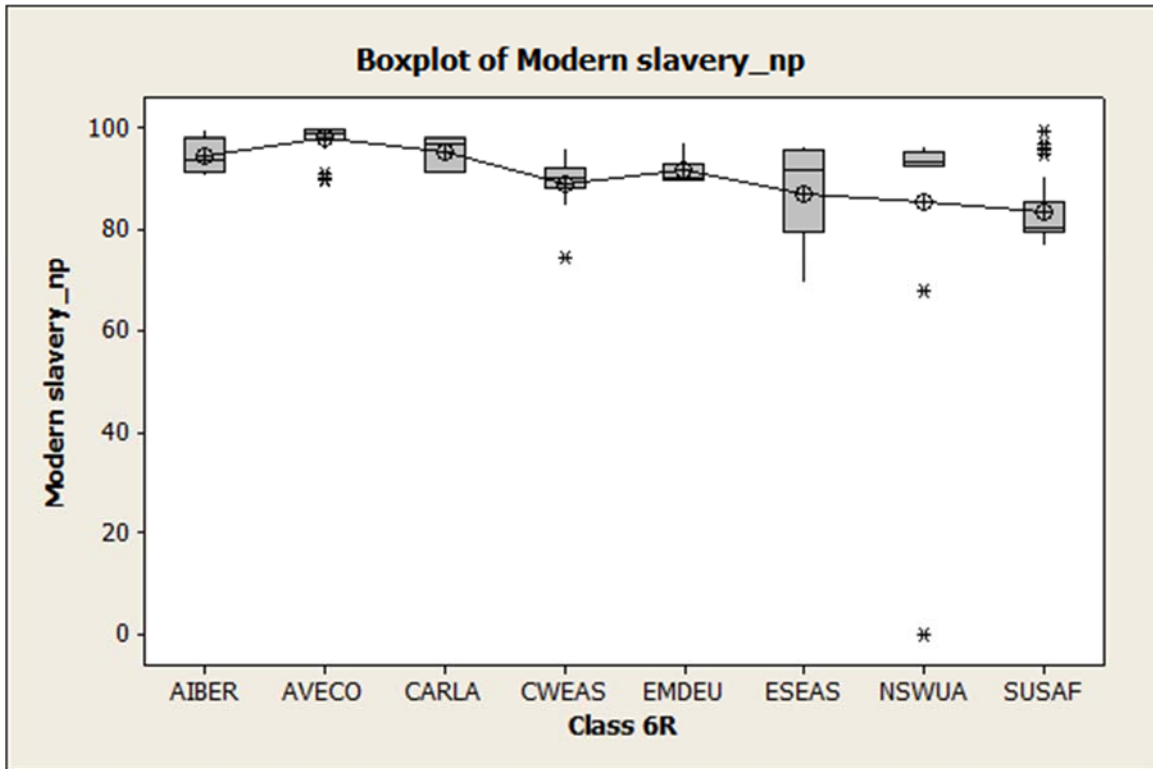
Source	DF	SS	MS	F	P
Class 6R	7	4022.0	574.6	5.97	0.000
Error	124	11925.0	96.2		
Total	131	15947.0			

S = 9.807 R-Sq = 25.22% R-Sq(adj) = 21.00%



Pooled StDev = 9.81

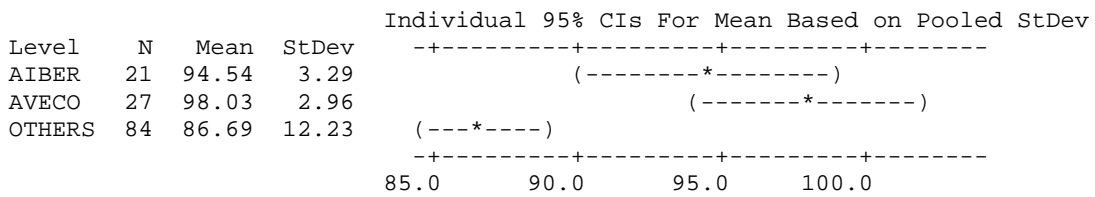
Boxplot of Modern slavery_np



One-way ANOVA: Modern slavery_np versus Class 3R

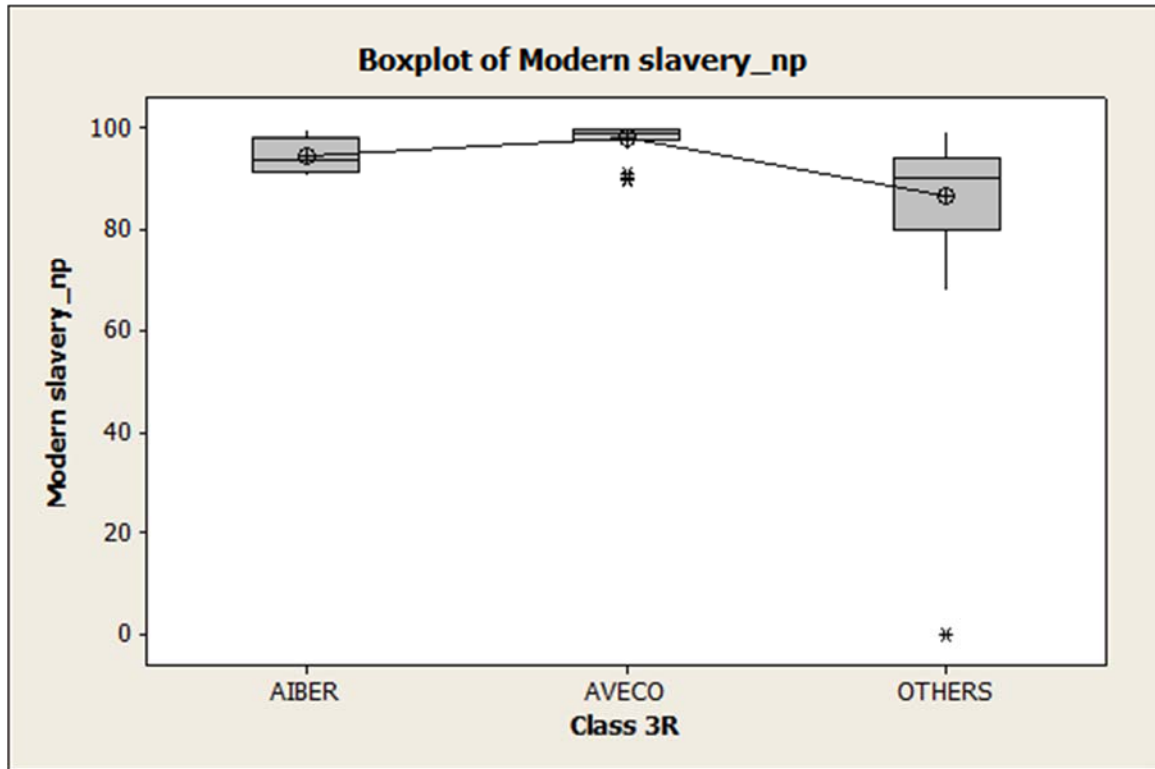
Source	DF	SS	MS	F	P
Class 3R	2	3088.6	1544.3	15.49	0.000
Error	129	12858.4	99.7		
Total	131	15947.0			

S = 9.984 R-Sq = 19.37% R-Sq(adj) = 18.12%



Pooled StDev = 9.98

Boxplot of Modern slavery_np



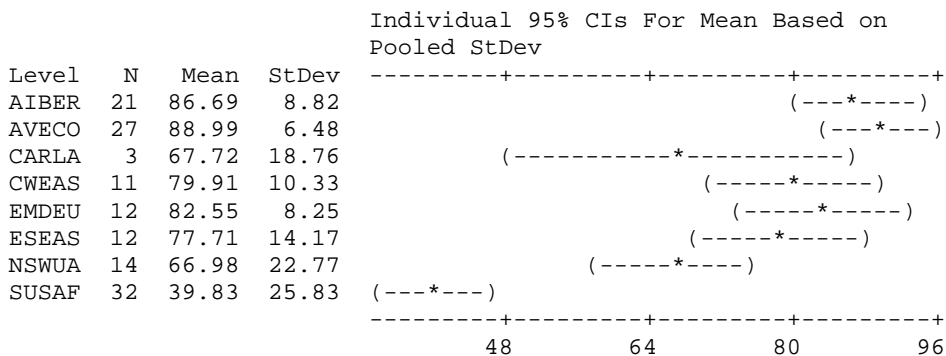
Obs: No agrupamento AVECO, vemos 3 *ouliers* que destoam do grupo, são eles: Eslovênia, Eslováquia e República Tcheca.

A Mauritânia lidera o ranking nesta categoria, com cerca de 155,600 indivíduos (4% da pop.), em situação de escravidão. fonte: <http://www.globalslaveryindex.org/country/mauritania/>

One-way ANOVA: Satisfied contraception (% of women) versus Class 6R

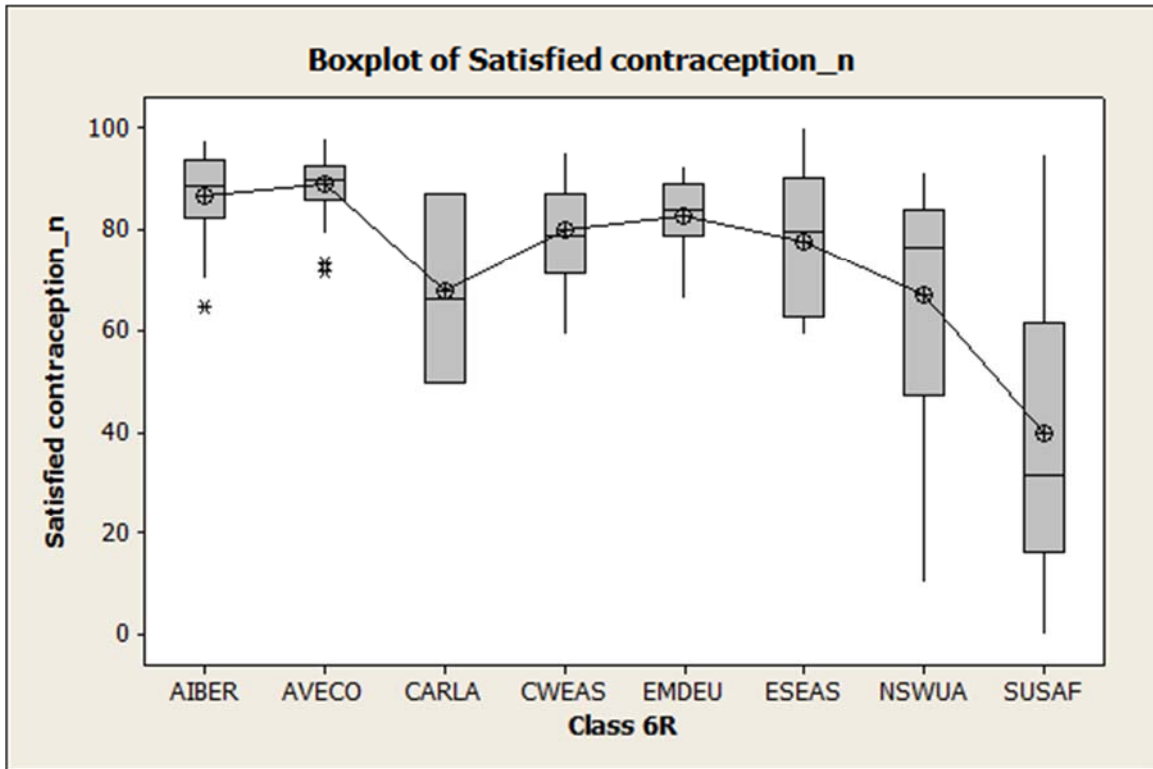
Source	DF	SS	MS	F	P
Class 6R	7	48235	6891	24.56	0.000
Error	124	34788	281		
Total	131	83022			

S = 16.75 R-Sq = 58.10% R-Sq(adj) = 55.73%



Pooled StDev = 16.75

Boxplot of Satisfied contraception (% of women)



One-way ANOVA: Satisfied contraception (% of women) versus Class 3R

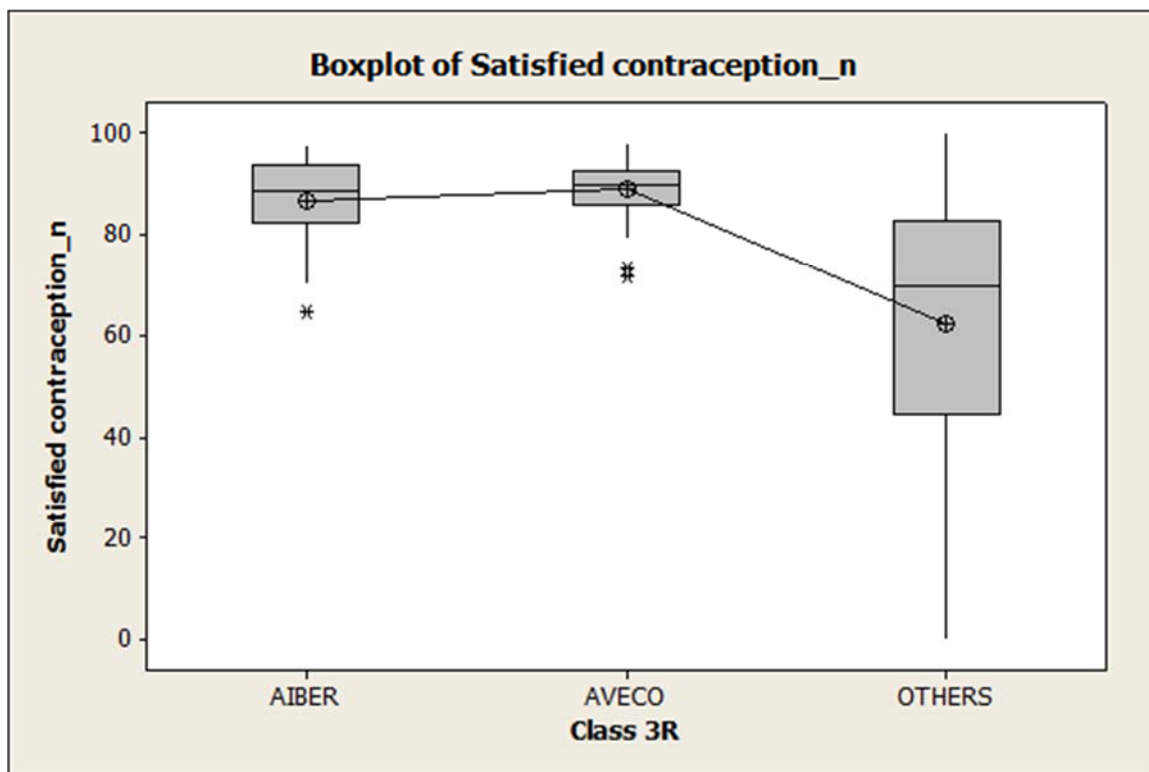
Source	DF	SS	MS	F	P
Class 3R	2	20501	10251	21.15	0.000
Error	129	62521	485		
Total	131	83022			

S = 22.01 R-Sq = 24.69% R-Sq(adj) = 23.53%

Level	N	Mean	StDev	Individual 95% CIs For Mean Based on Pooled StDev
AIBER	21	86.69	8.82	(-----*-----)
AVECO	27	88.99	6.48	(-----*-----)
OTHERS	84	62.11	26.86	(---*---)

Pooled StDev = 22.01

Boxplot of Satisfied contraception (% of women)



No agrupamento AIBER, figura como *outlier* a Guatemala, e no AVECO figuram Japão e Islândia respectivamente.

One-way ANOVA: Women treated with respect_n versus Class 6R

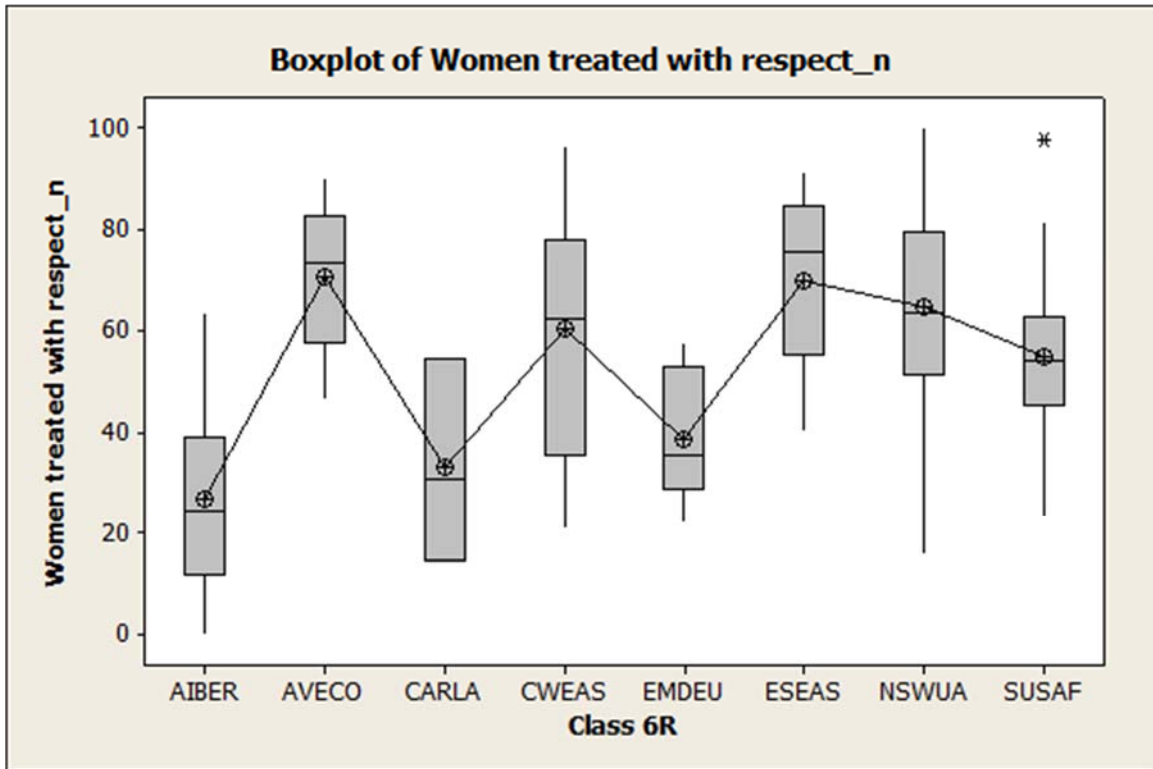
Source	DF	SS	MS	F	P
Class 6R	7	32034	4576	15.43	0.000
Error	124	36779	297		
Total	131	68813			

S = 17.22 R-Sq = 46.55% R-Sq(adj) = 43.53%

Level	N	Mean	StDev	Individual 95% CIs For Mean Based on Pooled StDev
AIBER	21	26.57	17.90	(--*--)
AVECO	27	70.33	13.59	(--*--)
CARLA	3	33.16	19.99	(-----*-----)
CWEAS	11	60.42	23.44	(-----*-----)
EMDEU	12	38.72	12.41	(-----*-----)
ESEAS	12	69.61	17.63	(-----*-----)
NSWUA	14	64.55	22.05	(-----*-----)
SUSAF	32	54.67	15.90	(--*--)

Pooled StDev = 17.22

Boxplot of Women treated with respect_n



One-way ANOVA: Women treated with respect_n versus Class 3R

Source	DF	SS	MS	F	P
Class 3R	2	23369	11685	33.17	0.000
Error	129	45444	352		
Total	131	68813			

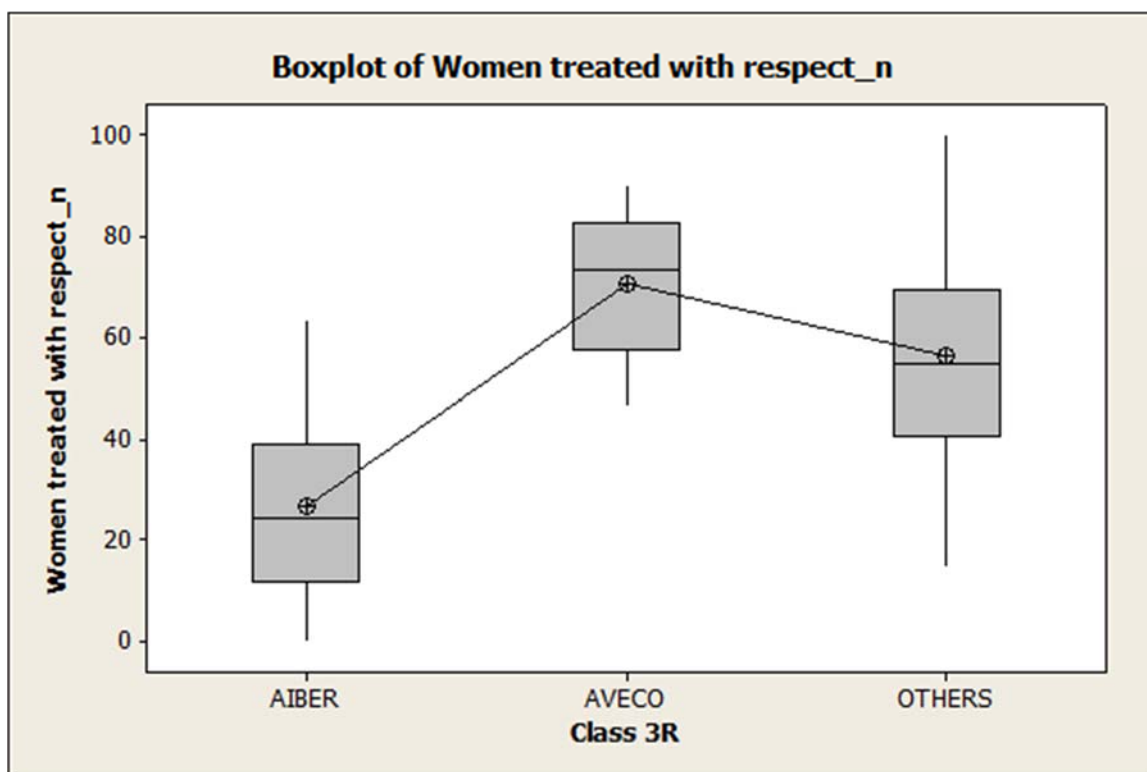
S = 18.77 R-Sq = 33.96% R-Sq(adj) = 32.94%

Level	N	Mean	StDev
AIBER	21	26.57	17.90
AVECO	27	70.33	13.59
OTHERS	84	56.16	20.31

Individual 95% CIs For Mean Based on Pooled StDev

Pooled StDev = 18.77

Boxplot of Women treated with respect_n



No agrupamento AIBER, os 3 países com os menores valores são: Perú com 24, República Dominicana com 18 e Colômbia com 17. O Brasil com 28, situa-se próximo destes.

One-way ANOVA: Tolerance for immigrants_n versus Class 6R

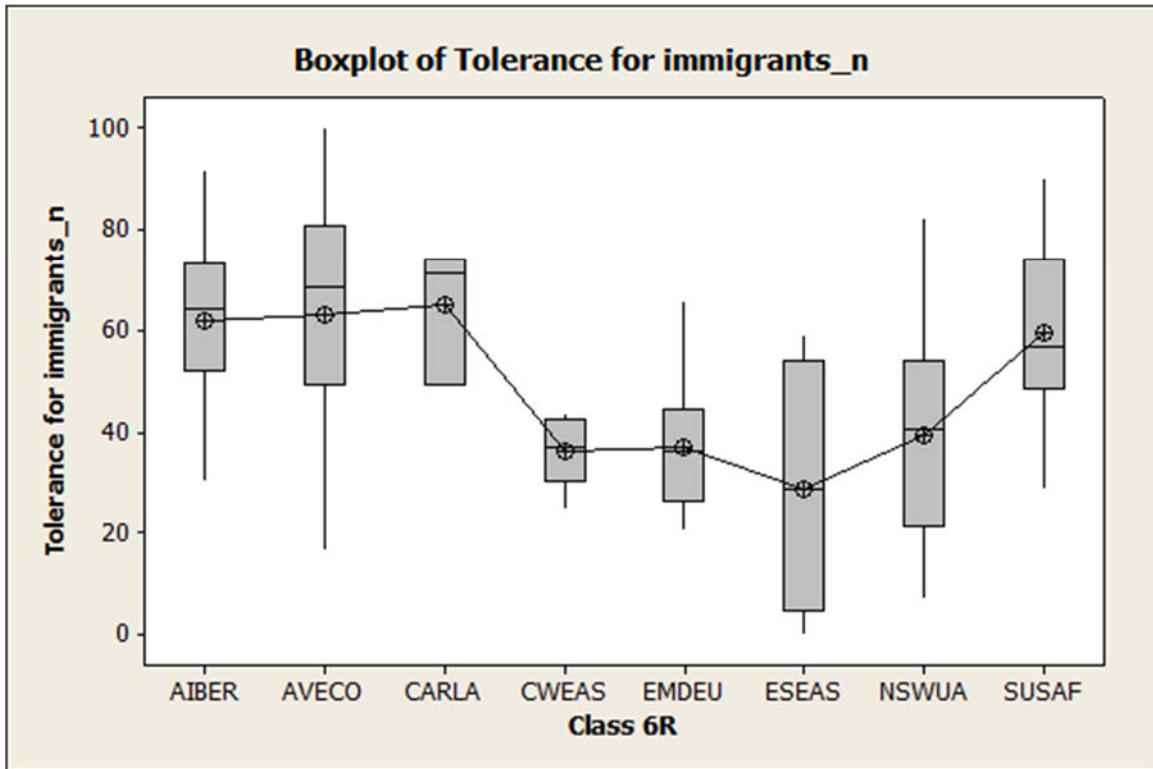
Source	DF	SS	MS	F	P
Class 6R	7	21882	3126	8.89	0.000
Error	124	43600	352		
Total	131	65483			

S = 18.75 R-Sq = 33.42% R-Sq(adj) = 29.66%

Level	N	Mean	StDev	Individual 95% CIs For Mean Based on Pooled StDev
AIBER	21	61.97	16.25	(---*---)
AVECO	27	62.86	24.27	(---*---)
CARLA	3	64.84	13.51	(-----*-----)
CWEAS	11	36.11	6.46	(-----*-----)
EMDEU	12	37.10	12.67	(-----*-----)
ESEAS	12	28.54	23.08	(-----*-----)
NSWUA	14	39.43	22.78	(-----*-----)
SUSAF	32	59.63	15.91	(---*---)

Pooled StDev = 18.75

Boxplot of Tolerance for immigrants_n



One-way ANOVA: Tolerance for immigrants_n versus Class 3R

Source	DF	SS	MS	F	P
Class 3R	2	8590	4295	9.74	0.000
Error	129	56893	441		
Total	131	65483			

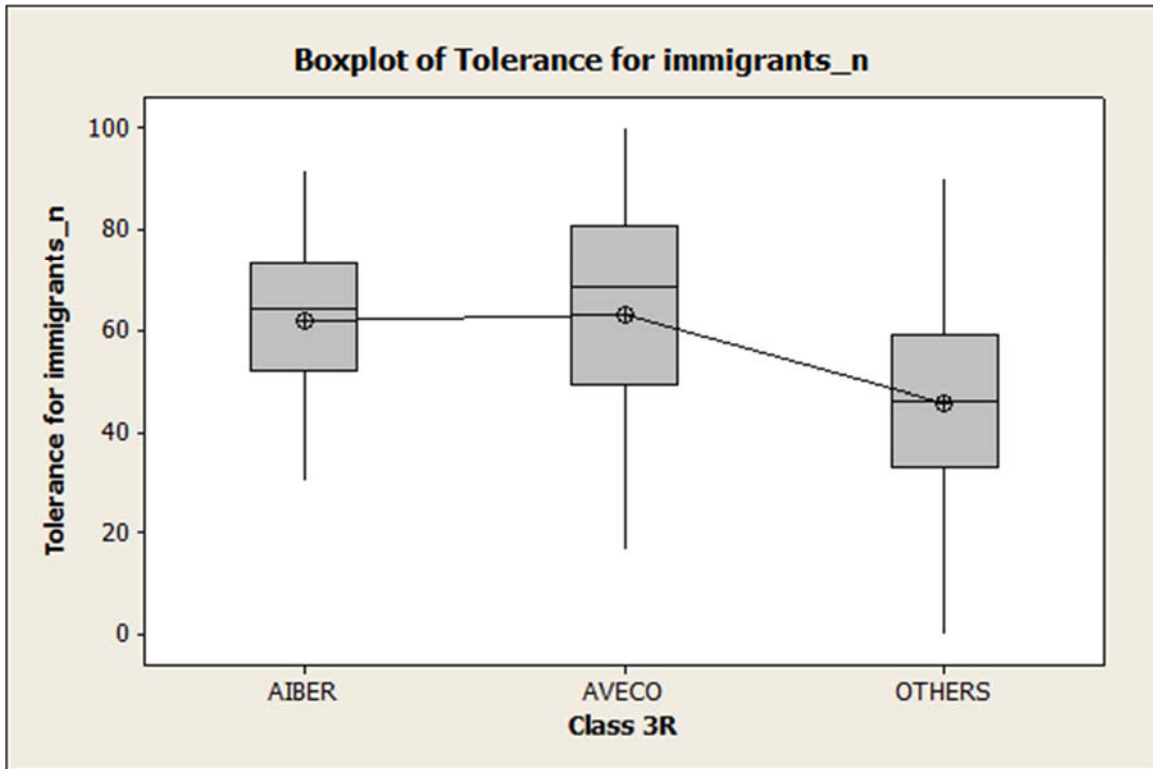
S = 21.00 R-Sq = 13.12% R-Sq(adj) = 11.77%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	CI Lower	CI Upper
AIBER	21	61.97	16.25	48.0	75.9
AVECO	27	62.86	24.27	56.0	71.7
OTHERS	84	45.71	20.91	24.8	66.6

Pooled StDev = 21.00

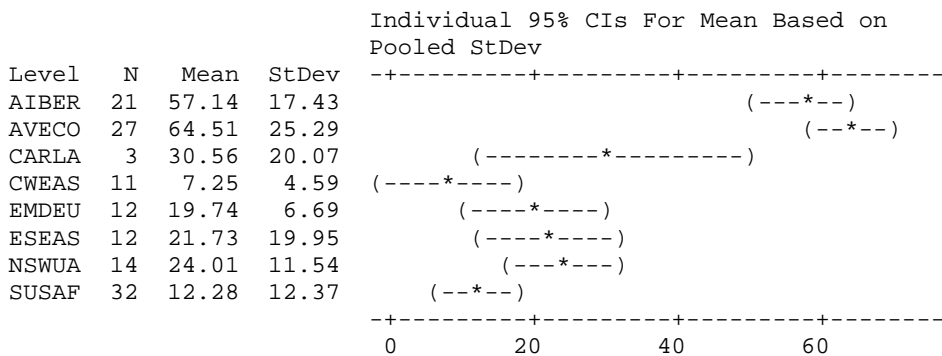
Boxplot of Tolerance for immigrants_n



One-way ANOVA: Tolerance for homosexuals_n versus Class 6R

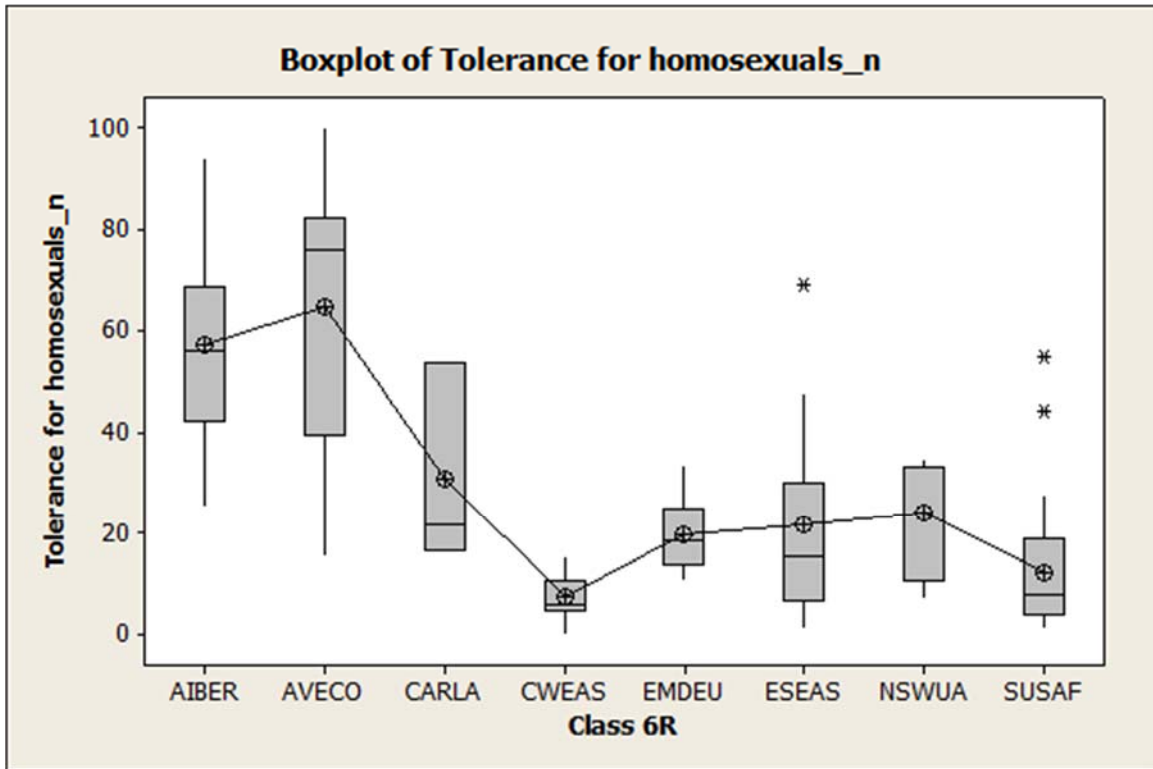
Source	DF	SS	MS	F	P
Class 6R	7	64859	9266	32.77	0.000
Error	124	35063	283		
Total	131	99923			

S = 16.82 R-Sq = 64.91% R-Sq(adj) = 62.93%



Pooled StDev = 16.82

Boxplot of Tolerance for homosexuals_n



One-way ANOVA: Tolerance for homosexuals_n versus Class 3R

Source	DF	SS	MS	F	P
Class 3R	2	61514	30757	103.30	0.000
Error	129	38409	298		
Total	131	99923			

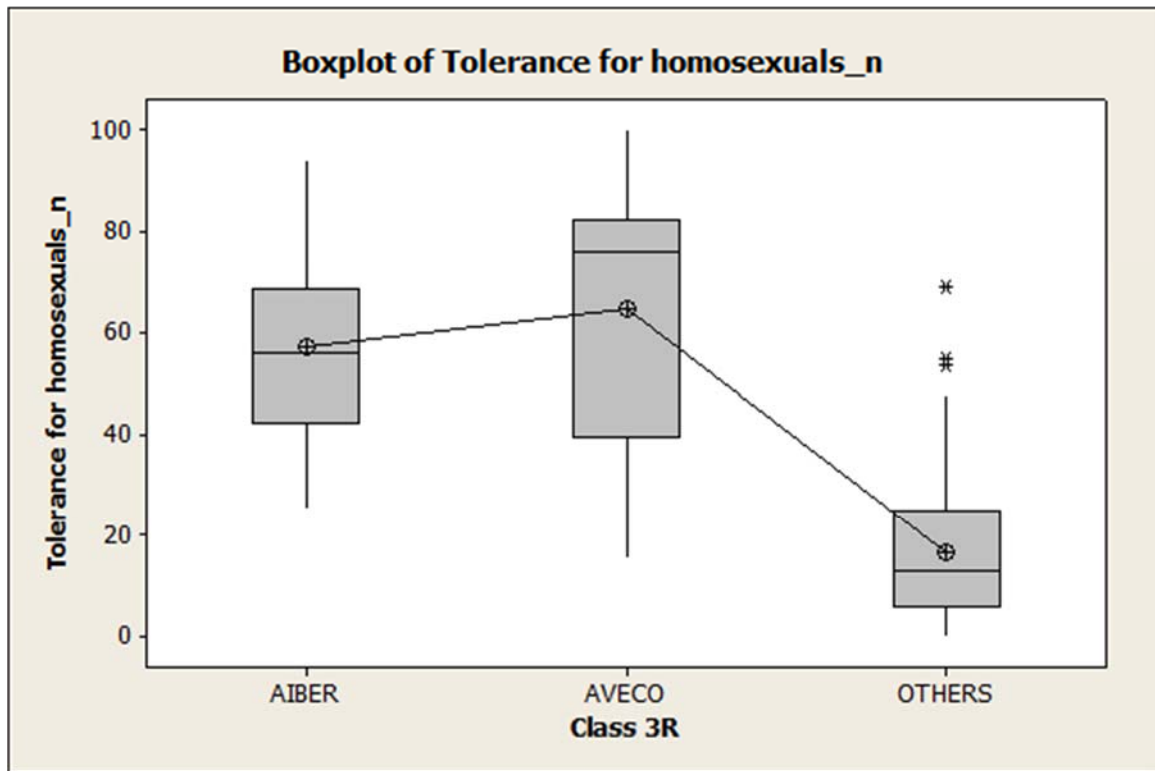
S = 17.26 R-Sq = 61.56% R-Sq(adj) = 60.97%

Level	N	Mean	StDev
AIBER	21	57.14	17.43
AVECO	27	64.51	25.29
OTHERS	84	16.64	13.76

Individual 95% CIs For Mean Based on Pooled StDev

Pooled StDev = 17.26

Boxplot of Tolerance for homosexuals_n



One-way ANOVA: Discrimination and violence against minorities_np versus Class 6R

Source	DF	SS	MS	F	P
Class 6R	7	24562	3509	10.61	0.000
Error	124	40999	331		
Total	131	65561			

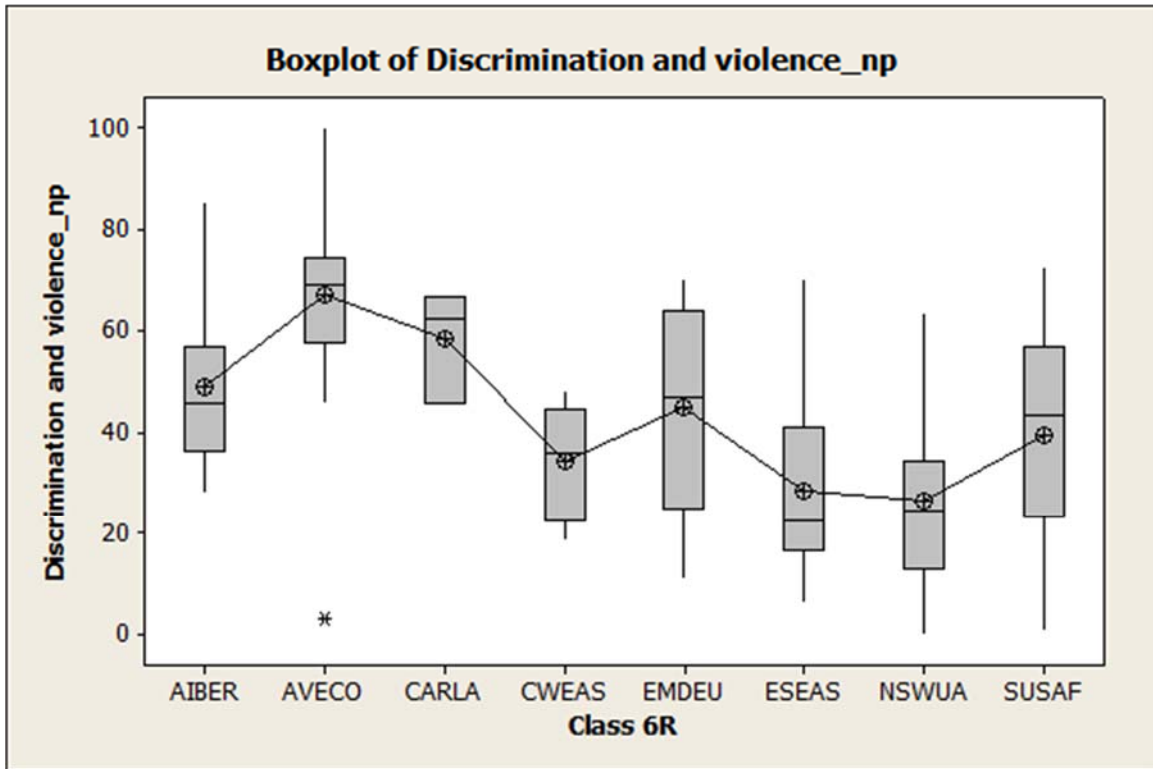
S = 18.18 R-Sq = 37.46% R-Sq(adj) = 33.93%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	CI Lower	CI Upper
AIBER	21	48.96	16.08	16.00	81.92
AVECO	27	67.10	19.37	27.36	106.84
CARLA	3	58.27	11.10	36.07	80.47
CWEAS	11	34.26	10.61	13.04	55.48
EMDEU	12	44.86	20.39	4.08	85.64
ESEAS	12	28.24	17.60	0.00	56.48
NSWUA	14	26.21	18.64	0.00	52.42
SUSAF	32	39.55	19.84	0.00	79.10

Pooled StDev = 18.18

Boxplot of Discrimination and violence against minorities_np



One-way ANOVA: Discrimination and violence against minorities_np versus Class 3R

Source	DF	SS	MS	F	P
Class 3R	2	19650	9825	27.61	0.000
Error	129	45911	356		
Total	131	65561			

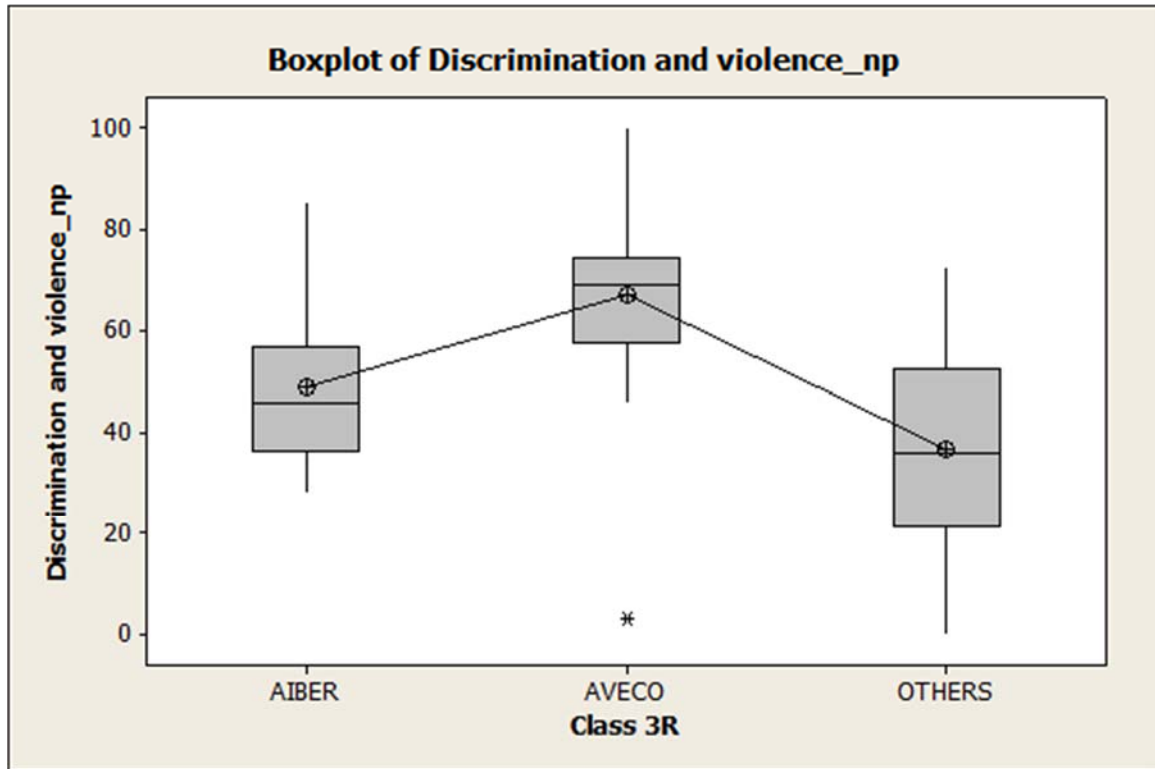
S = 18.87 R-Sq = 29.97% R-Sq(adj) = 28.89%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	CI Lower	CI Upper
AIBER	21	48.96	16.08	36.00	61.92
AVECO	27	67.10	19.37	47.76	86.44
OTHERS	84	36.45	19.32	17.13	55.77

Pooled StDev = 18.87

Boxplot of Discrimination and violence against minorities_np



Obs: No agrupamento AVECO, Israel aparece como *outlier* destoando significativamente de seu grupo.

One-way ANOVA: Religious tolerance_n versus Class 6R

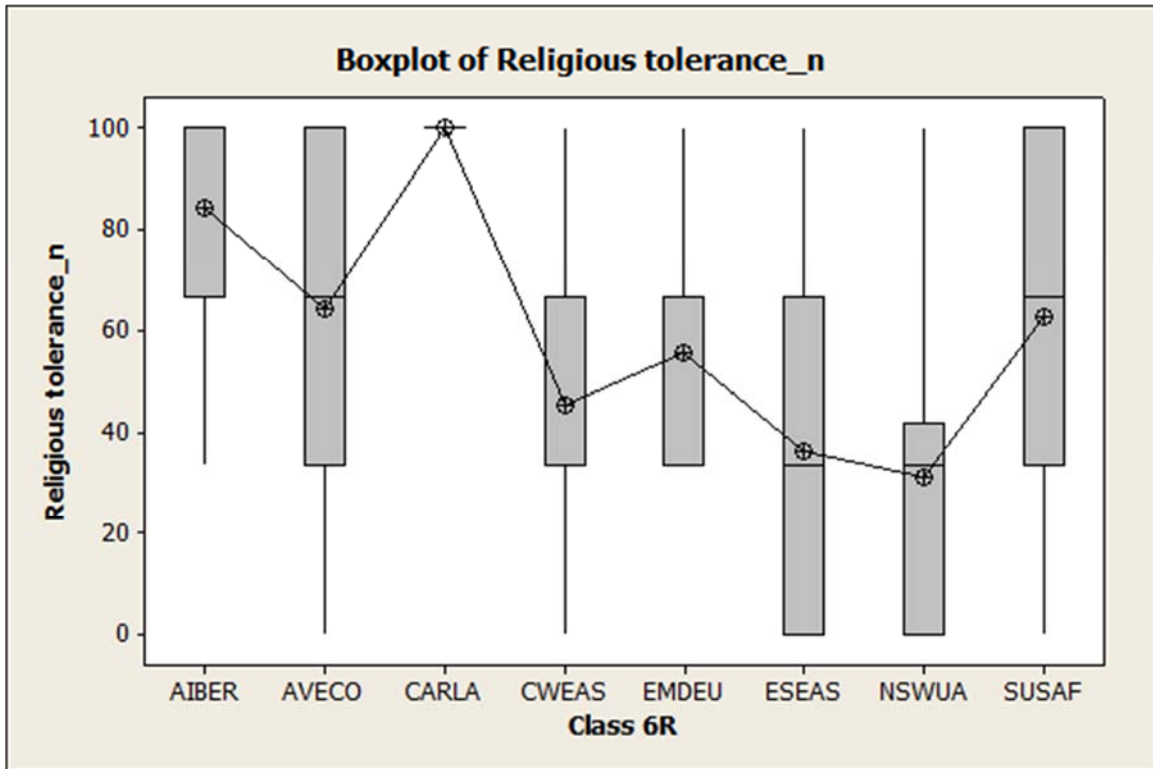
Source	DF	SS	MS	F	P
Class 6R	7	38868	5553	6.76	0.000
Error	124	101830	821		
Total	131	140699			

S = 28.66 R-Sq = 27.63% R-Sq(adj) = 23.54%

Level	N	Mean	StDev	Individual 95% CIs For Mean Based on Pooled StDev
AIBER	21	84.13	22.65	(---*---)
AVECO	27	64.20	27.62	(--*--)
CARLA	3	100.00	0.00	(-----*-----)
CWEAS	11	45.45	26.97	(-----*-----)
EMDEU	12	55.56	21.71	(-----*-----)
ESEAS	12	36.11	38.82	(----*----)
NSWUA	14	30.95	30.56	(----*----)
SUSAF	32	62.50	31.40	(---*---)

Pooled StDev = 28.66

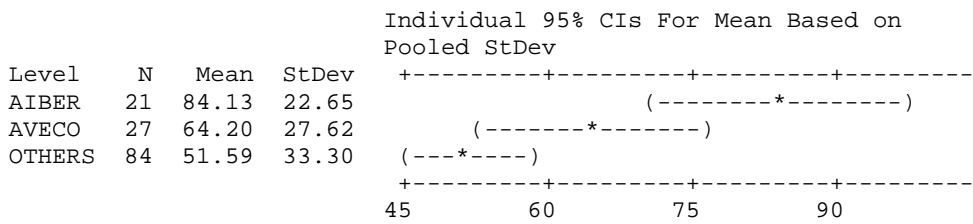
Boxplot of Religious tolerance_n



One-way ANOVA: Religious tolerance_n versus Class 3R

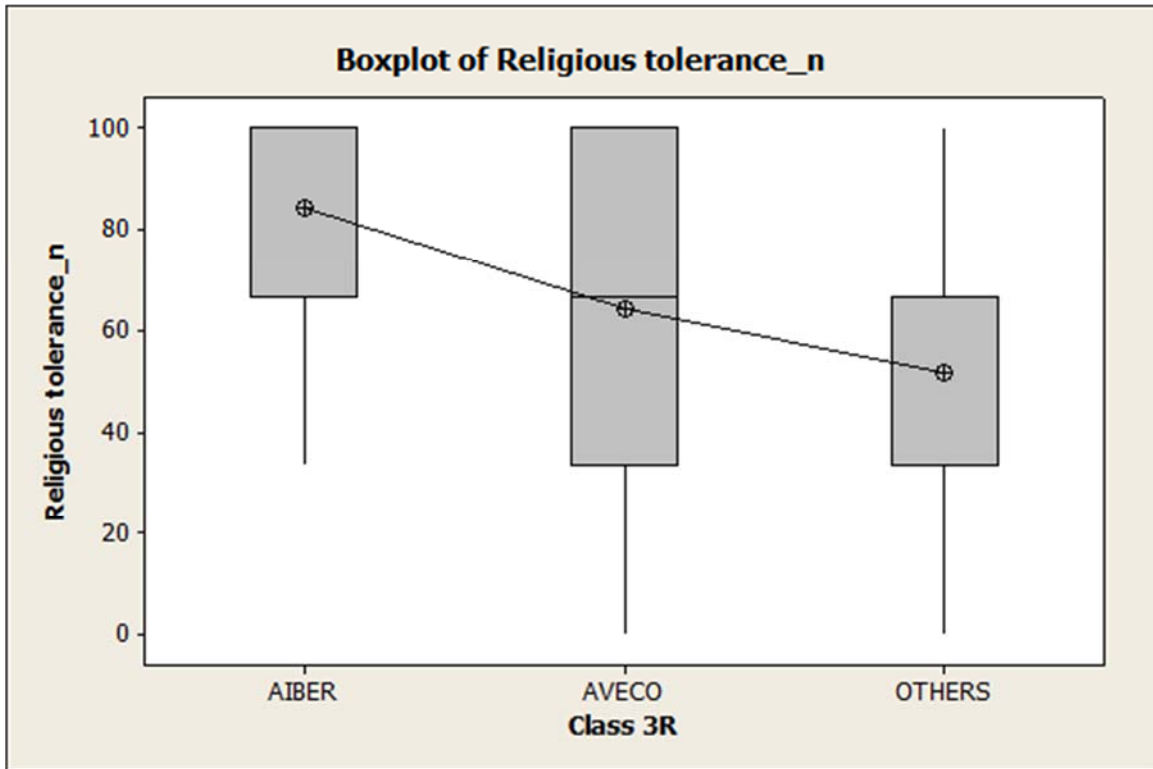
Source	DF	SS	MS	F	P
Class 3R	2	18588	9294	9.82	0.000
Error	129	122111	947		
Total	131	140699			

S = 30.77 R-Sq = 13.21% R-Sq(adj) = 11.87%



Pooled StDev = 30.77

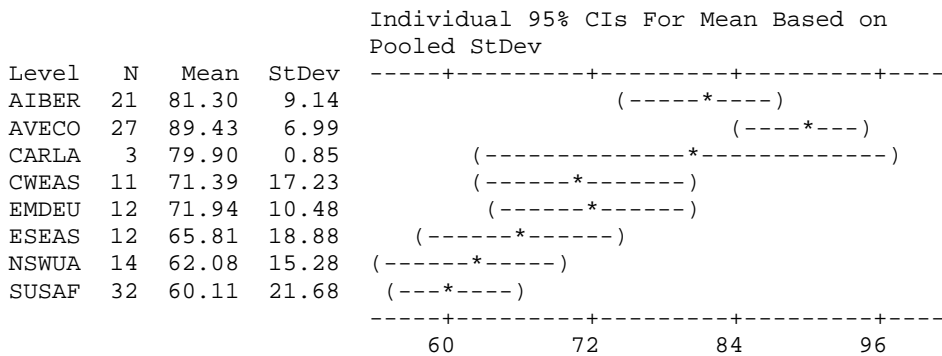
Boxplot of Religious tolerance_n



One-way ANOVA: Community safety net_n versus Class 6R

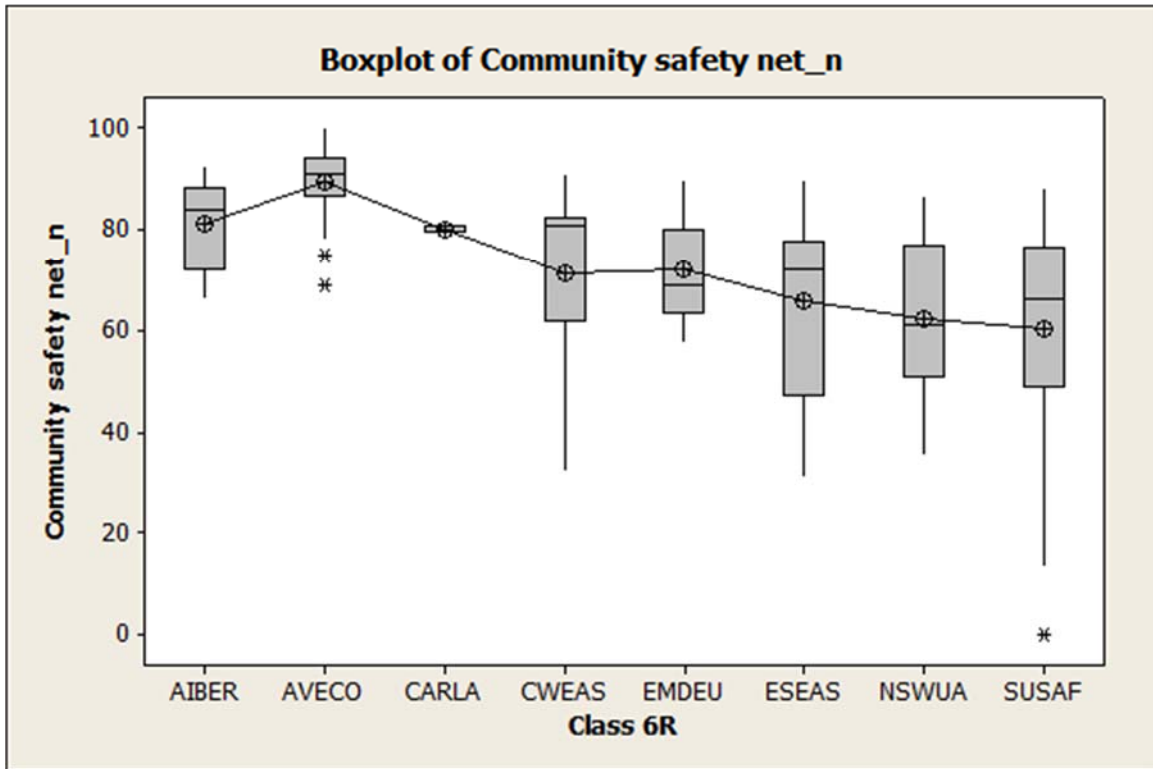
Source	DF	SS	MS	F	P
Class 6R	7	16517	2360	10.21	0.000
Error	124	28652	231		
Total	131	45169			

S = 15.20 R-Sq = 36.57% R-Sq(adj) = 32.99%



Pooled StDev = 15.20

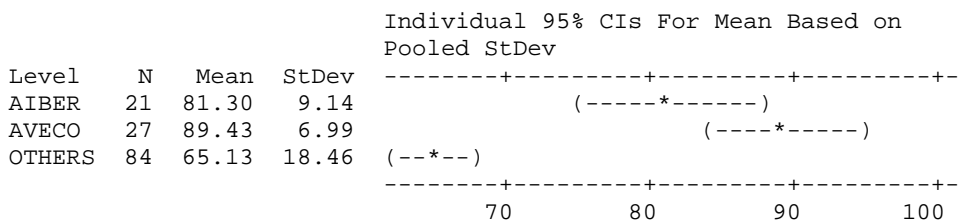
Boxplot of Community safety net_n



One-way ANOVA: Community safety net_n versus Class 3R

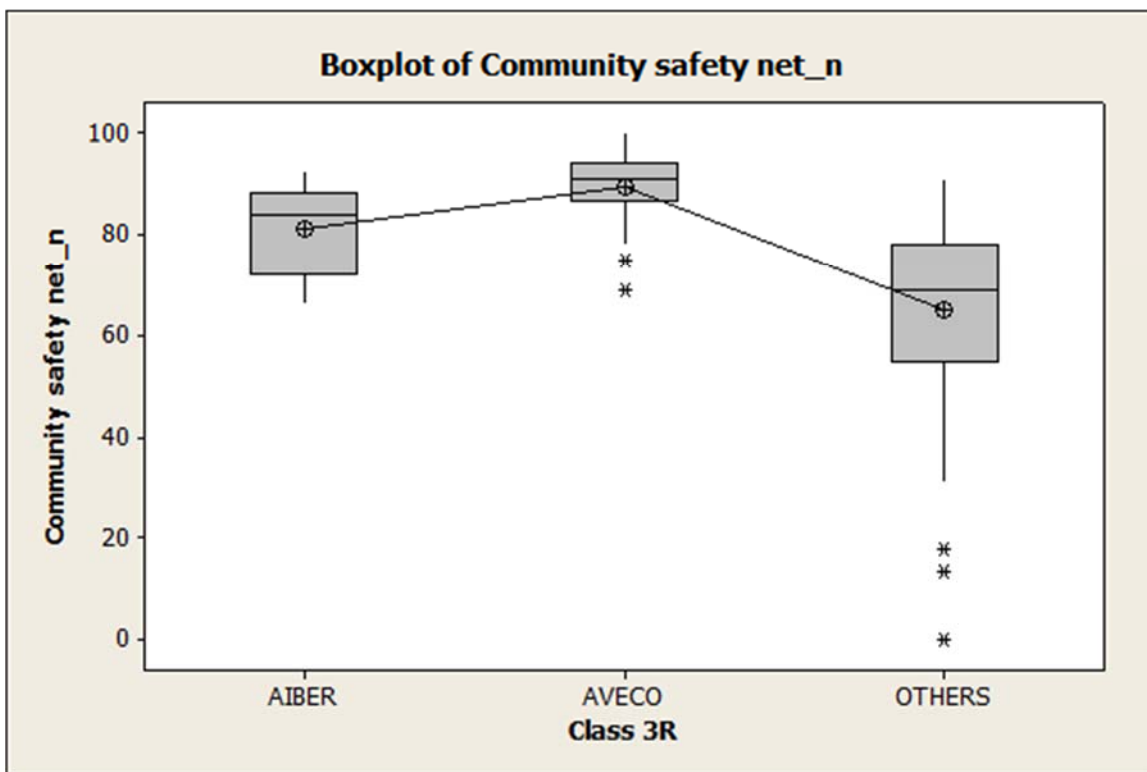
Source	DF	SS	MS	F	P
Class 3R	2	13933	6966	28.77	0.000
Error	129	31236	242		
Total	131	45169			

S = 15.56 R-Sq = 30.85% R-Sq(adj) = 29.77%



Pooled StDev = 15.56

Boxplot of Community safety net_n



Obs: Os 2 outliers que aparecem no agrupamento AVECO, são: Grécia e Coreia do Sul (Republic of Korea).

One-way ANOVA: Years of tertiary schooling_n versus Class 6R

Source	DF	SS	MS	F	P
Class 6R	7	27592	3942	17.54	0.000
Error	124	27861	225		
Total	131	55453			

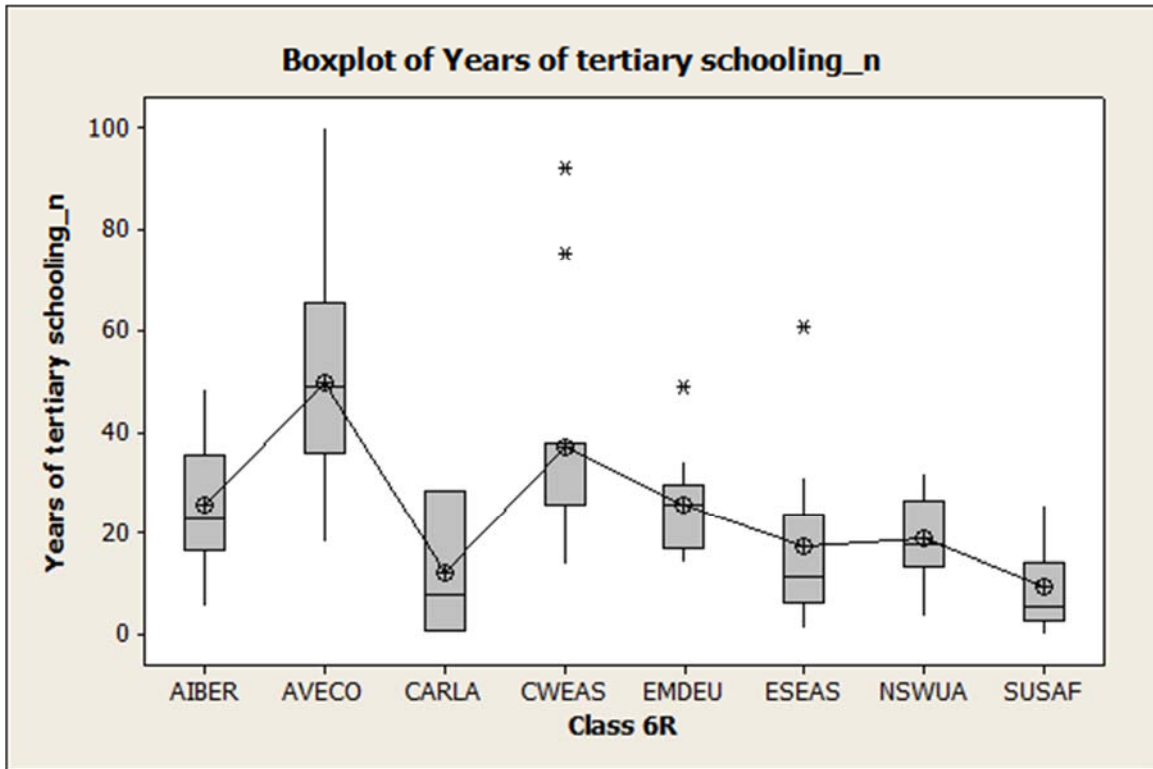
S = 14.99 R-Sq = 49.76% R-Sq(adj) = 46.92%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev
AIBER	21	25.74	12.61
AVECO	27	49.63	20.36
CARLA	3	12.16	14.36
CWEAS	11	37.06	24.20
EMDEU	12	25.36	9.60
ESEAS	12	17.40	16.31
NSWUA	14	18.86	8.65
SUSAF	32	9.23	9.37

Pooled StDev = 14.99

Boxplot of Years of tertiary schooling_n



Years of tertiary schooling = 15 missing values

One-way ANOVA: Years of tertiary schooling_n versus Class 3R

Source	DF	SS	MS	F	P
Class 3R	2	20370	10185	37.45	0.000
Error	129	35082	272		
Total	131	55453			

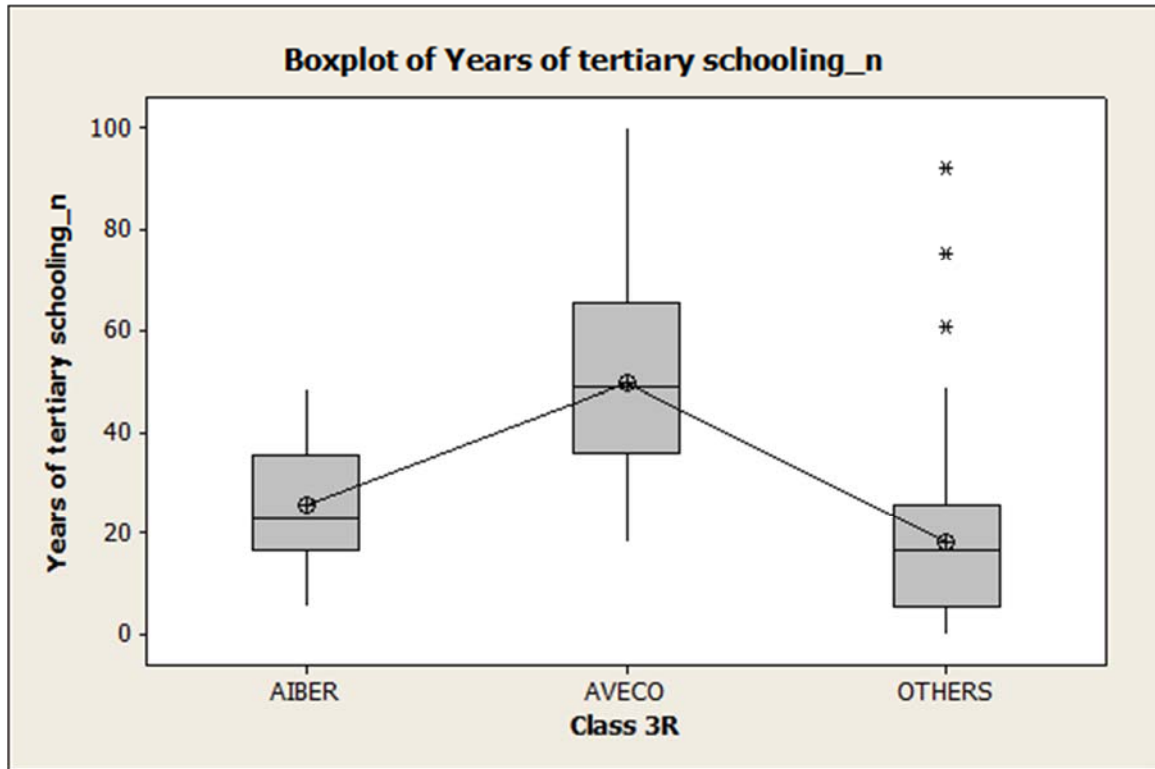
S = 16.49 R-Sq = 36.73% R-Sq(adj) = 35.75%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	CI Lower	CI Upper
AIBER	21	25.74	12.61	10.00	41.48
AVECO	27	49.63	20.36	10.00	89.26
OTHERS	84	18.05	15.95	0.00	36.10

Pooled StDev = 16.49

Boxplot of Years of tertiary schooling_n

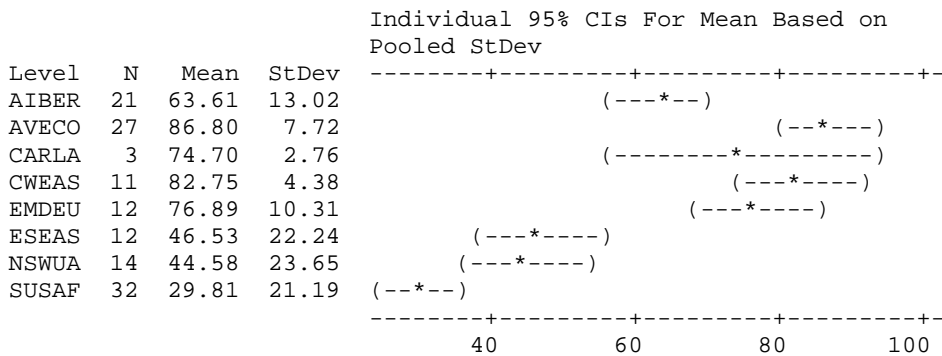


Years of tertiary schooling = 15 missing values

One-way ANOVA: Women`s years in school_n versus Class 6R

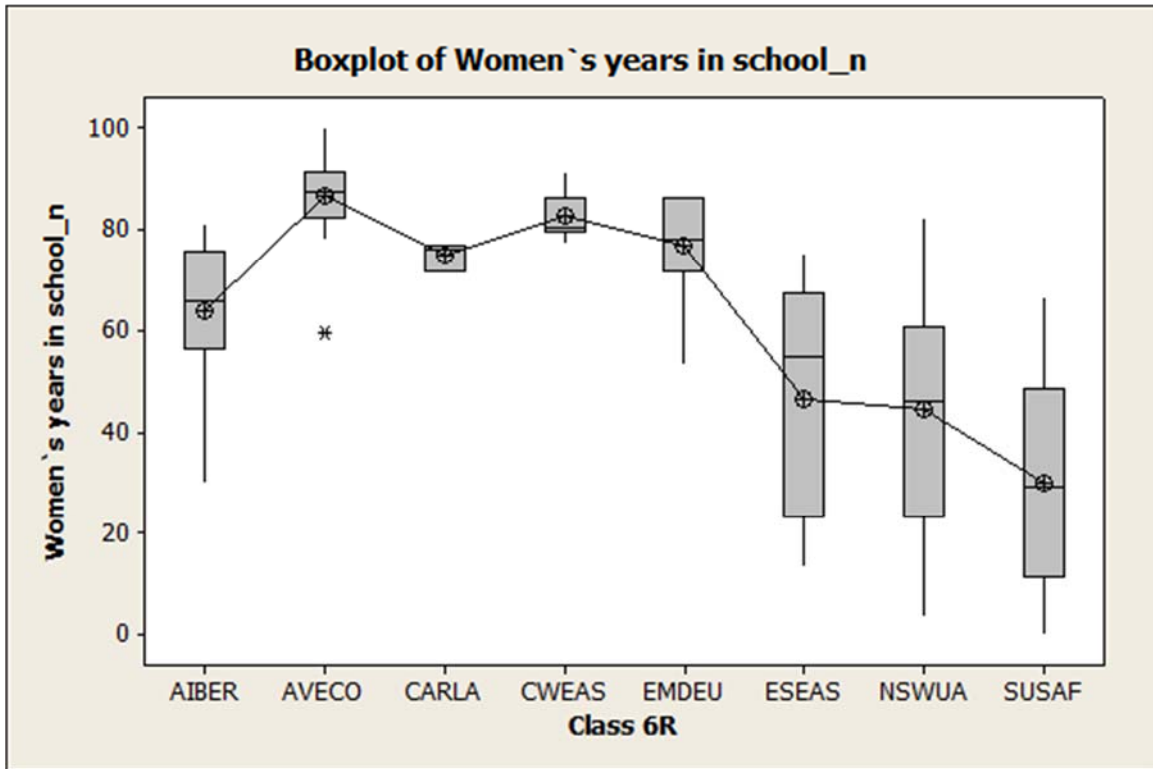
Source	DF	SS	MS	F	P
Class 6R	7	64076	9154	34.46	0.000
Error	124	32939	266		
Total	131	97015			

S = 16.30 R-Sq = 66.05% R-Sq(adj) = 64.13%



Pooled StDev = 16.30

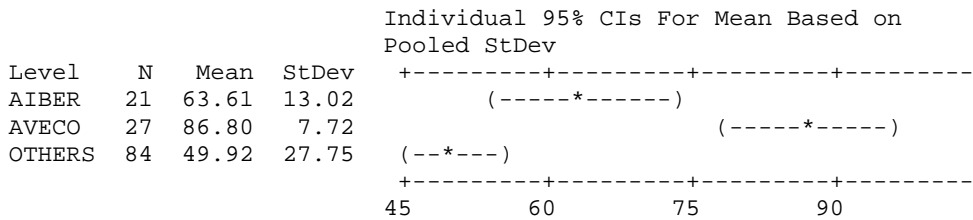
Boxplot of Women`s years in school_n



One-way ANOVA: Women`s years in school_n versus Class 3R

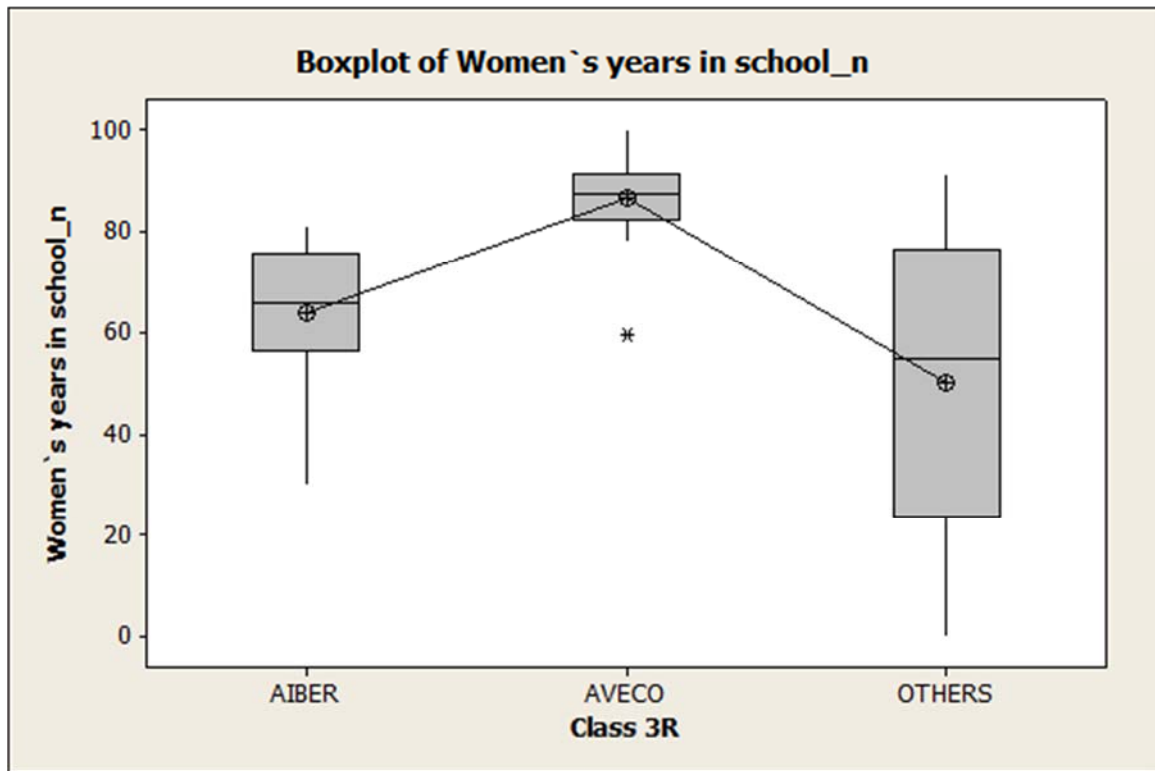
Source	DF	SS	MS	F	P
Class 3R	2	28180	14090	26.41	0.000
Error	129	68835	534		
Total	131	97015			

S = 23.10 R-Sq = 29.05% R-Sq(adj) = 27.95%



Pooled StDev = 23.10

Boxplot of Women`s years in school_n

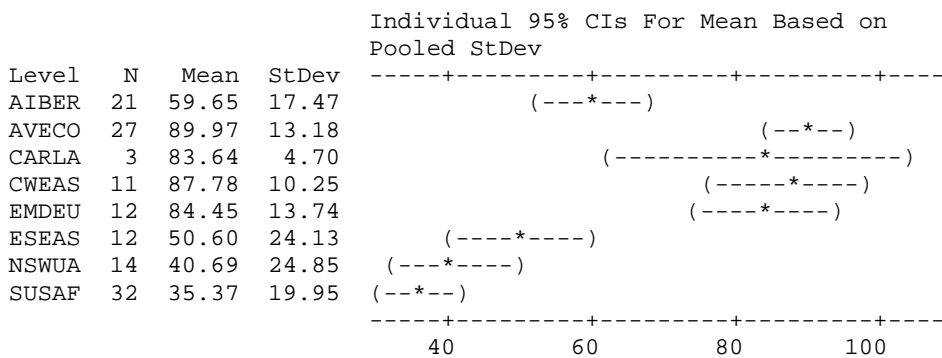


Obs: No agrupamento AVECO, a Islândia figura como *outlier* .

One-way ANOVA: Inequality in the attainment of education versus Class 6R

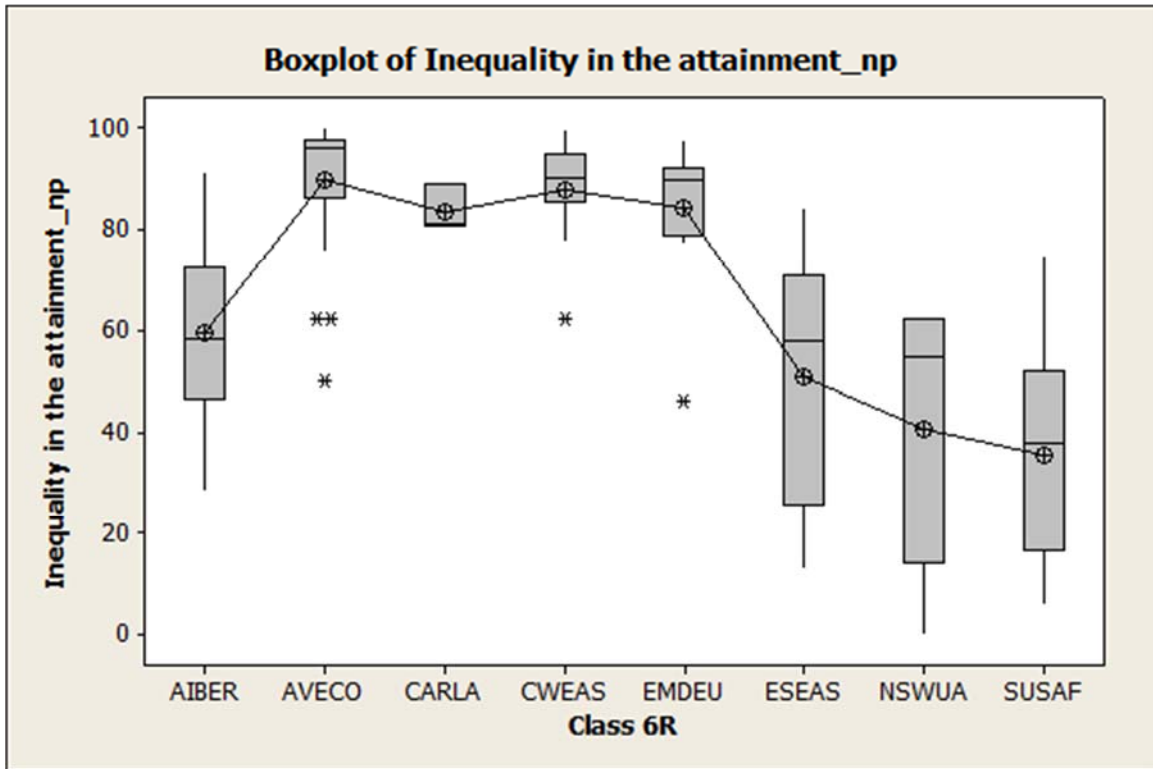
Source	DF	SS	MS	F	P
Class 6R	7	66599	9514	29.09	0.000
Error	124	40562	327		
Total	131	107161			

S = 18.09 R-Sq = 62.15% R-Sq(adj) = 60.01%



Pooled StDev = 18.09

Boxplot of Inequality in the attainment of education



One-way ANOVA: Inequality in the attainment of education versus Class 3R

Source	DF	SS	MS	F	P
Class 3R	2	26558	13279	21.25	0.000
Error	129	80603	625		
Total	131	107161			

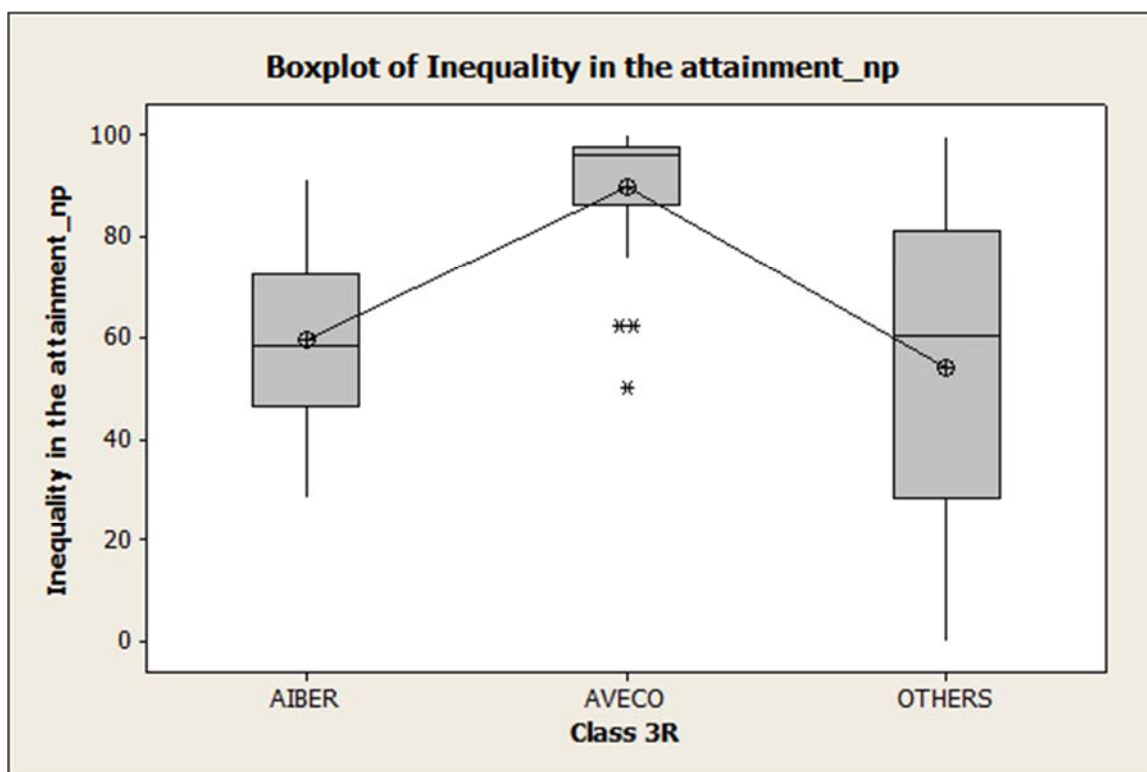
S = 25.00 R-Sq = 24.78% R-Sq(adj) = 23.62%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	CI Lower	CI Upper
AIBER	21	59.65	17.47	42.18	77.12
AVECO	27	89.97	13.18	76.79	103.15
OTHERS	84	54.03	29.04	25.00	83.06

Pooled StDev = 25.00

Boxplot of Inequality in the attainment of education



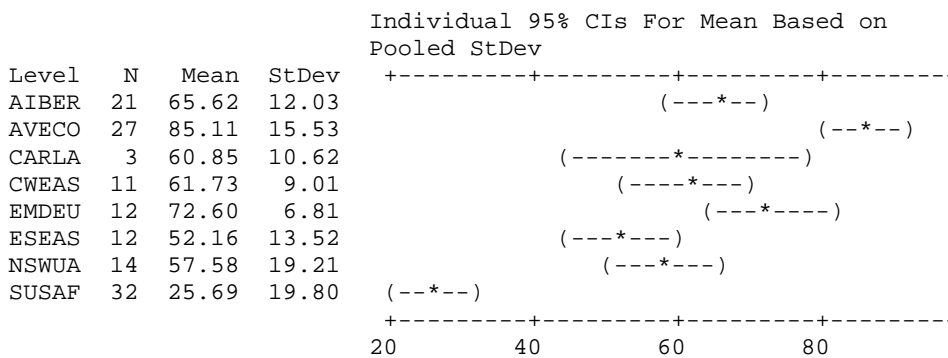
Obs: Japão, Nova Zelândia e Coreia do Sul

Inequality in the attainment of education = 15 missing values

One-way ANOVA: IDH-2013_n versus Class 6R

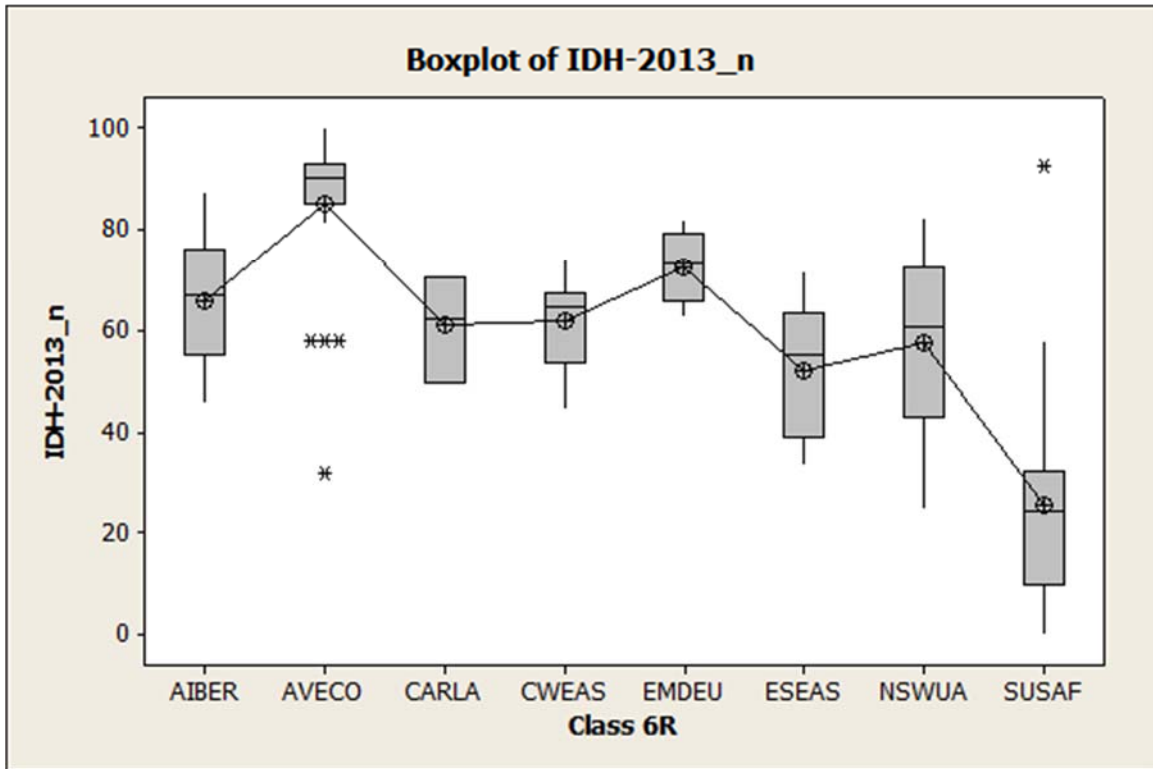
Source	DF	SS	MS	F	P
Class 6R	7	57615	8231	34.39	0.000
Error	124	29677	239		
Total	131	87292			

S = 15.47 R-Sq = 66.00% R-Sq(adj) = 64.08%



Pooled StDev = 15.47

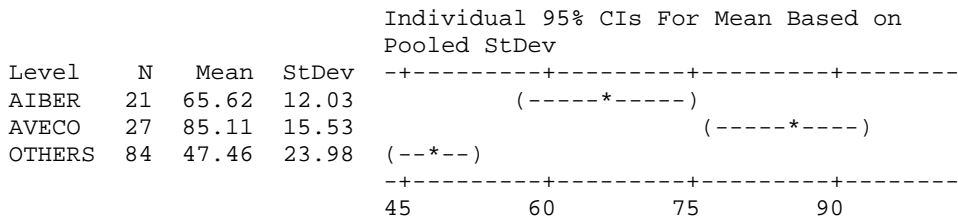
Boxplot of IDH-2013_n



One-way ANOVA: IDH-2013_n versus Class 3R

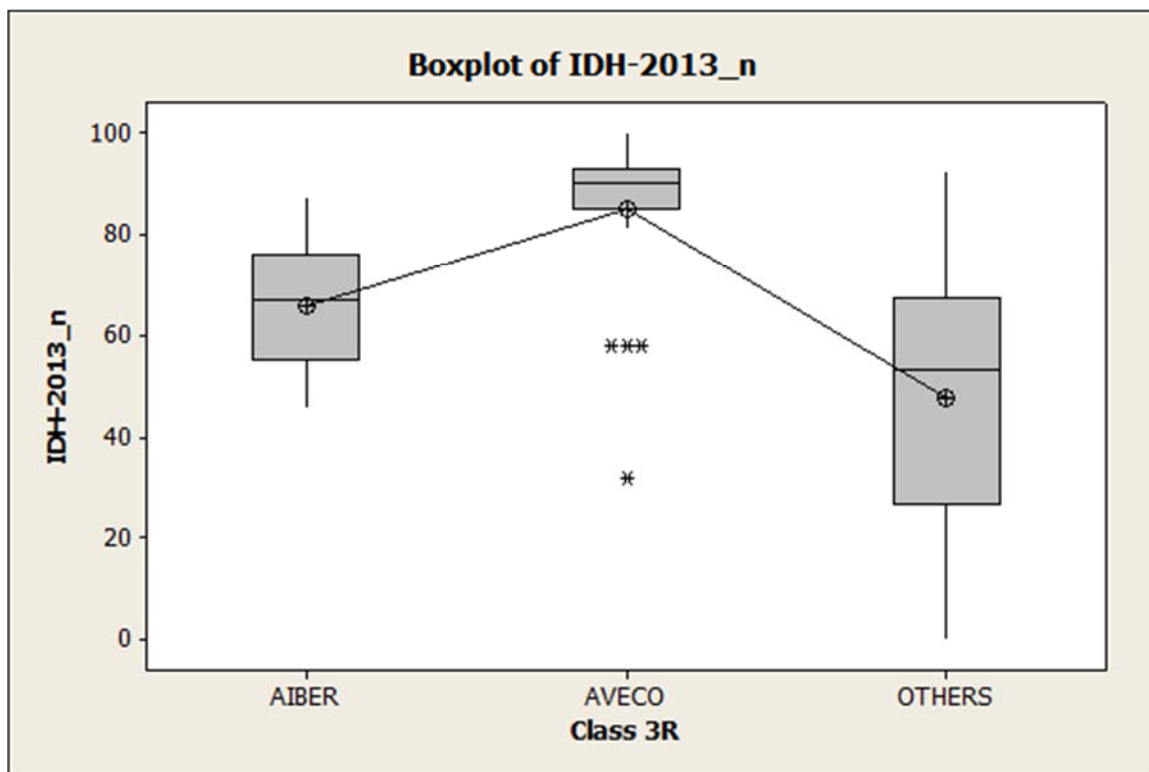
Source	DF	SS	MS	F	P
Class 3R	2	30392	15196	34.45	0.000
Error	129	56900	441		
Total	131	87292			

S = 21.00 R-Sq = 34.82% R-Sq(adj) = 33.81%



Pooled StDev = 21.00

Boxplot of IDH-2013_n



Obs: Coreia do Sul (Republic of Korea), Letônia (Latvia), Holanda e Suíça figura como outliers no gráfico acima.

IDH-2013 = 7 missing values

*Coreia, Holanda e Letônia estavam sem valores na planilha. E Suíça com 0.53 na planilha original

One-way ANOVA: Governança_n versus Class 6R

Source	DF	SS	MS	F	P
Class 6R	7	56636	8091	35.45	0.000
Error	124	28301	228		
Total	131	84938			

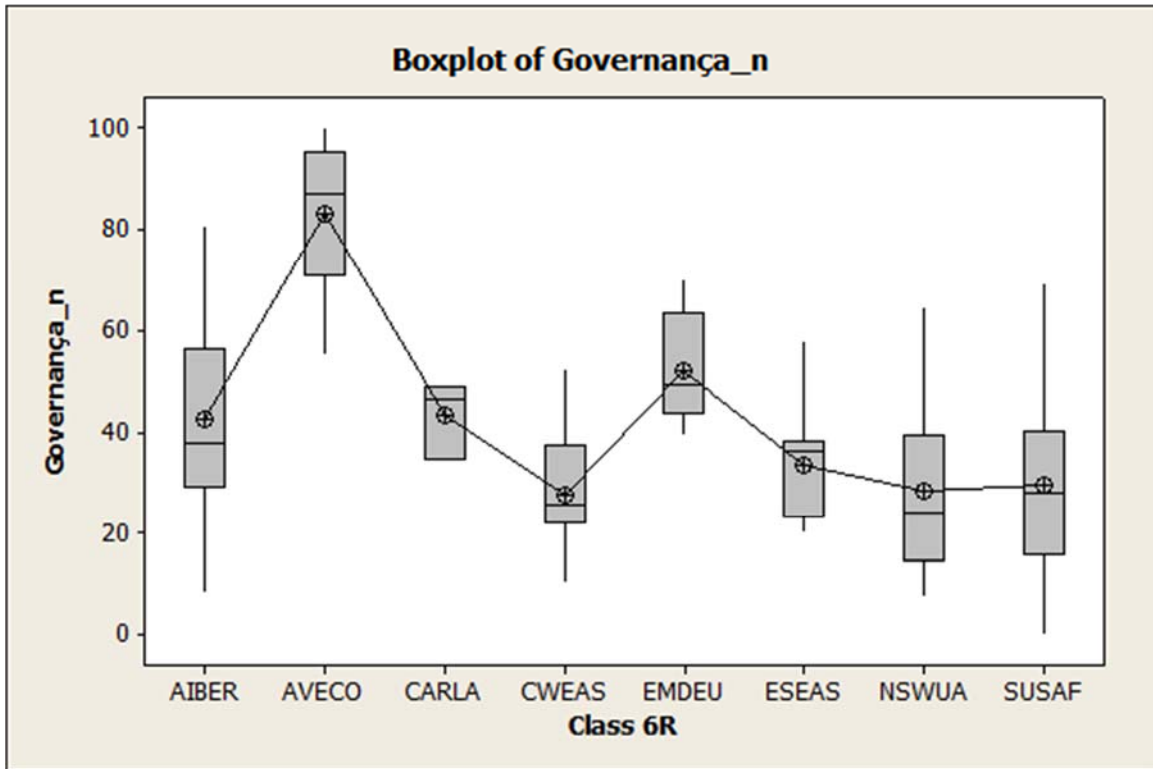
S = 15.11 R-Sq = 66.68% R-Sq(adj) = 64.80%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	CI Lower	CI Upper
AIBER	21	42.48	18.82	23.66	61.30
AVECO	27	83.25	13.38	69.87	96.63
CARLA	3	43.25	7.61	35.64	50.86
CWEAS	11	27.38	12.38	14.00	40.76
EMDEU	12	51.93	11.21	40.72	63.14
ESEAS	12	33.61	10.47	23.14	44.08
NSWUA	14	28.46	16.21	12.25	44.67
SUSAF	32	29.58	16.95	12.63	46.53

Pooled StDev = 15.11

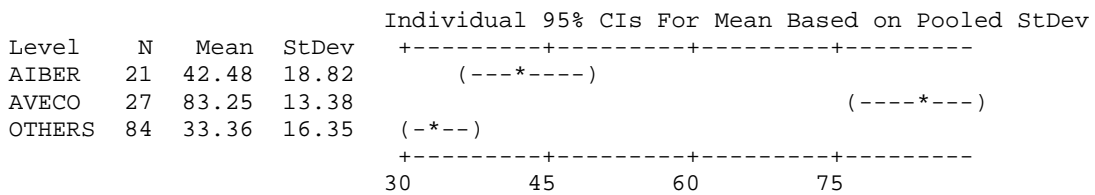
Boxplot of Governança_n



One-way ANOVA: Governança_n versus Class 3R

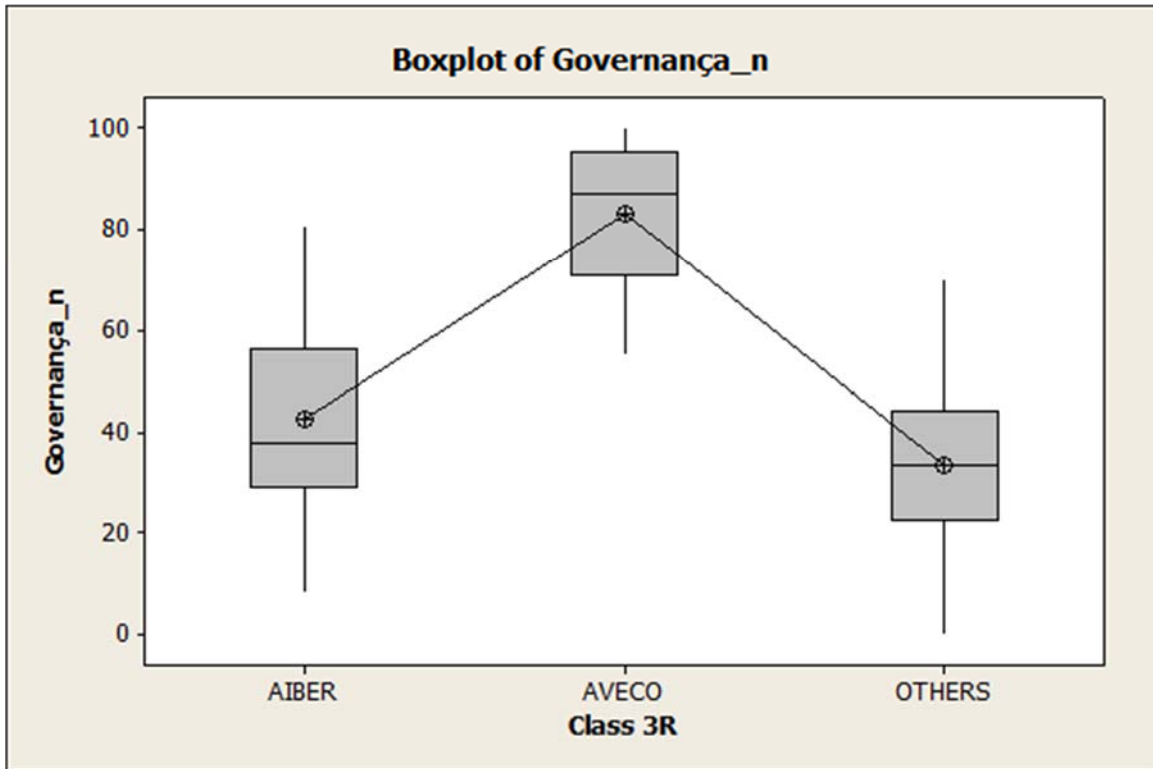
Source	DF	SS	MS	F	P
Class 3R	2	51019	25510	97.02	0.000
Error	129	33919	263		
Total	131	84938			

S = 16.22 R-Sq = 60.07% R-Sq(adj) = 59.45%



Pooled StDev = 16.22

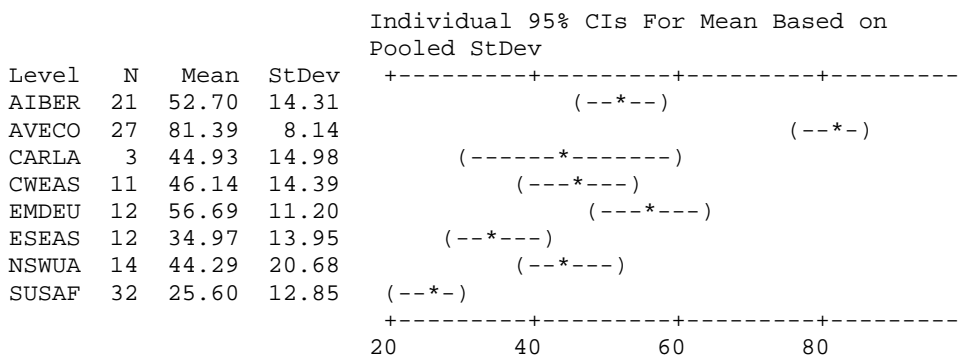
Boxplot of Governança_n



One-way ANOVA: EPI Score_n versus Class 6R

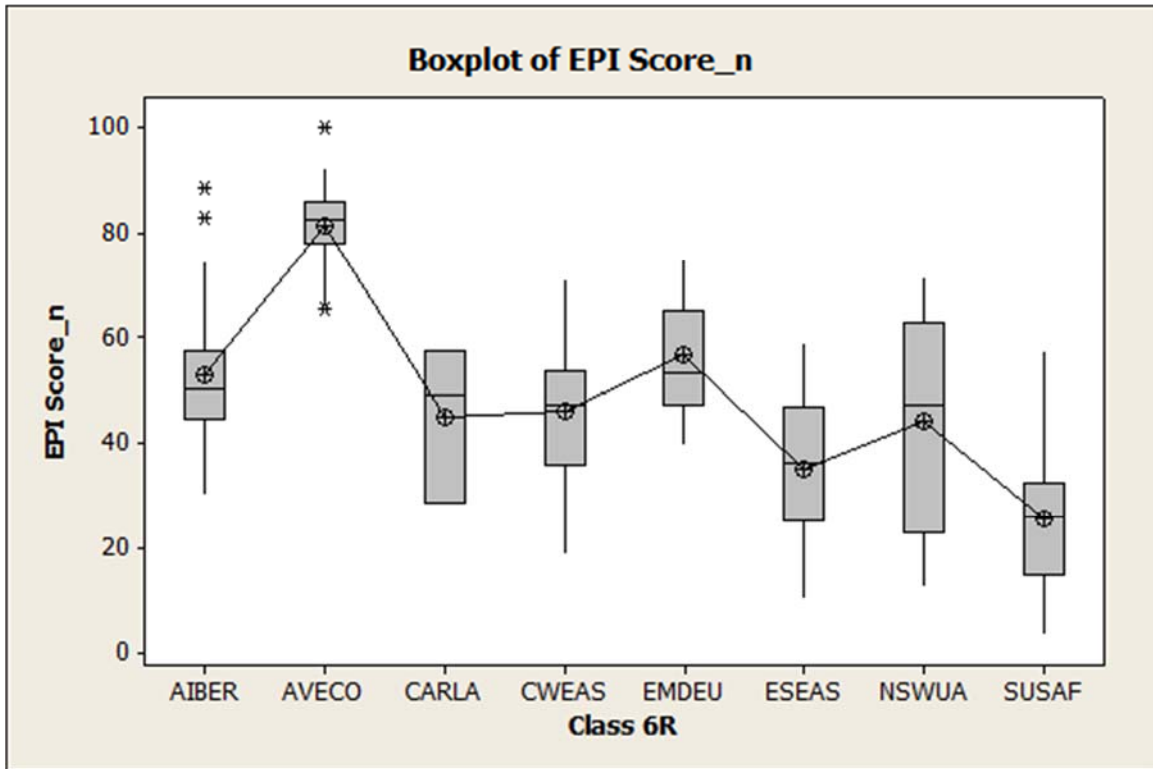
Source	DF	SS	MS	F	P
Class 6R	7	49668	7095	39.04	0.000
Error	124	22535	182		
Total	131	72203			

S = 13.48 R-Sq = 68.79% R-Sq(adj) = 67.03%



Pooled StDev = 13.48

Boxplot of EPI Score_n



One-way ANOVA: EPI Score_n versus Class 3R

Source	DF	SS	MS	F	P
Class 3R	2	39017	19508	75.83	0.000
Error	129	33186	257		
Total	131	72203			

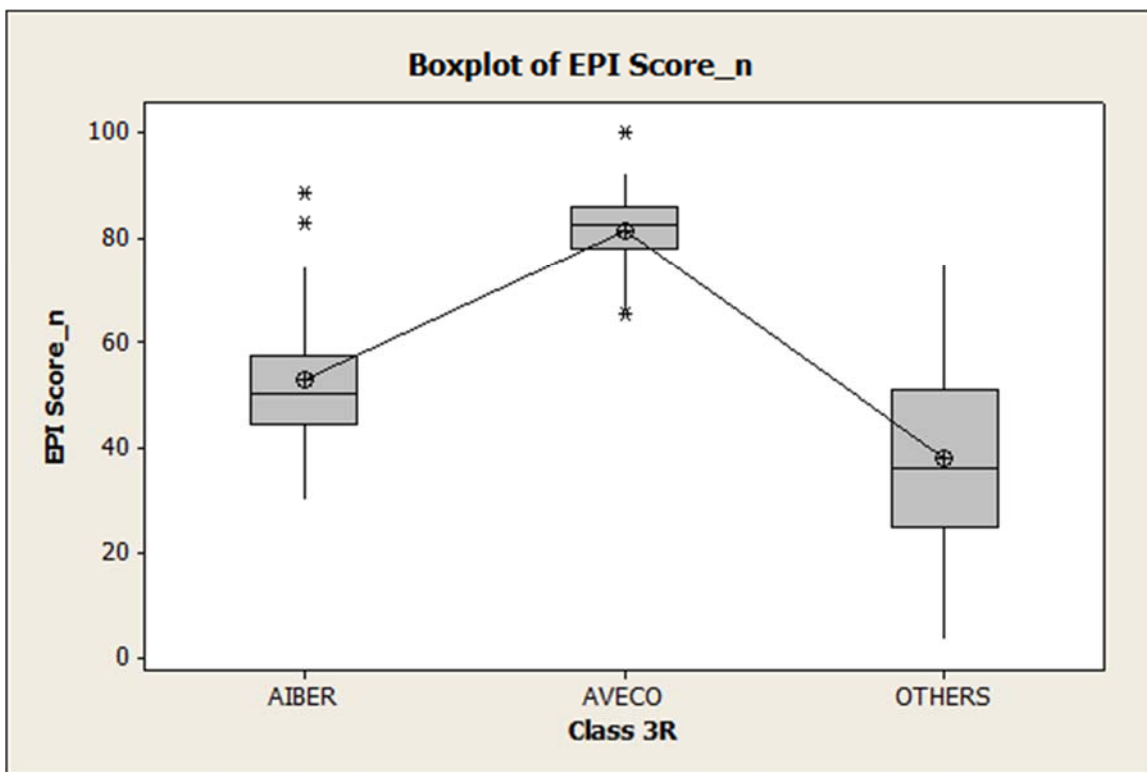
S = 16.04 R-Sq = 54.04% R-Sq(adj) = 53.33%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev
AIBER	21	52.70	14.31
AVECO	27	81.39	8.14
OTHERS	84	37.87	18.16

Pooled StDev = 16.04

Boxplot of EPI Score_n



Obs: Destacaram-se positivamente como *outliers* os países Portugal e Espanha no agrupamento AIBER e Suíça no AVECO. E negativamente como *outlier* a Coreia do Sul (Republic of Korea).

One-way ANOVA: EV-Water Resources_n versus Class 6R

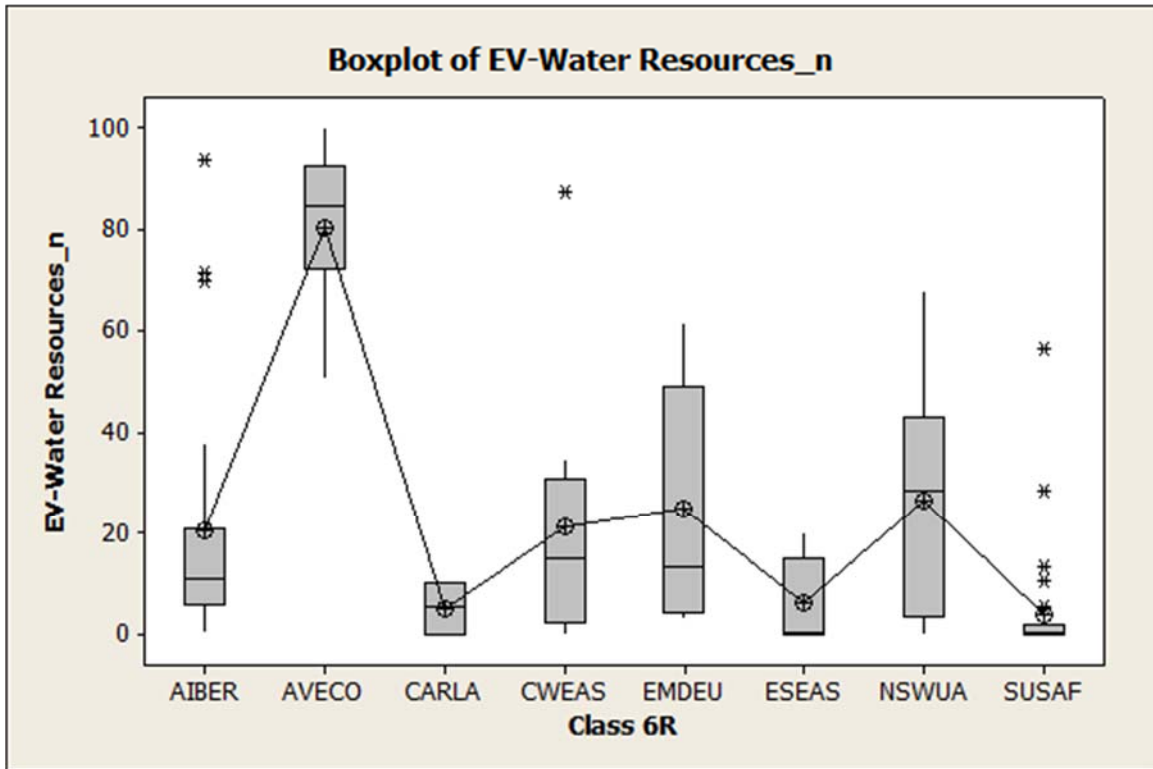
Source	DF	SS	MS	F	P
Class 6R	7	101460	14494	42.74	0.000
Error	124	42049	339		
Total	131	143509			

S = 18.41 R-Sq = 70.70% R-Sq(adj) = 69.05%

Level	N	Mean	StDev	Individual 95% CIs For Mean Based on Pooled StDev
AIBER	21	20.39	26.00	(---*)
AVECO	27	80.29	14.88	(---*)
CARLA	3	5.14	5.06	(-----*-----)
CWEAS	11	21.11	25.13	(---*---)
EMDEU	12	24.74	23.16	(---*---)
ESEAS	12	6.21	8.14	(---*---)
NSWUA	14	26.19	21.49	(---*---)
SUSAF	32	3.99	11.05	(---*)

Pooled StDev = 18.41

Boxplot of EV-Water Resources_n



One-way ANOVA: EV-Water Resources_n versus Class 3R

Source	DF	SS	MS	F	P
Class 3R	2	93314	46657	119.91	0.000
Error	129	50195	389		
Total	131	143509			

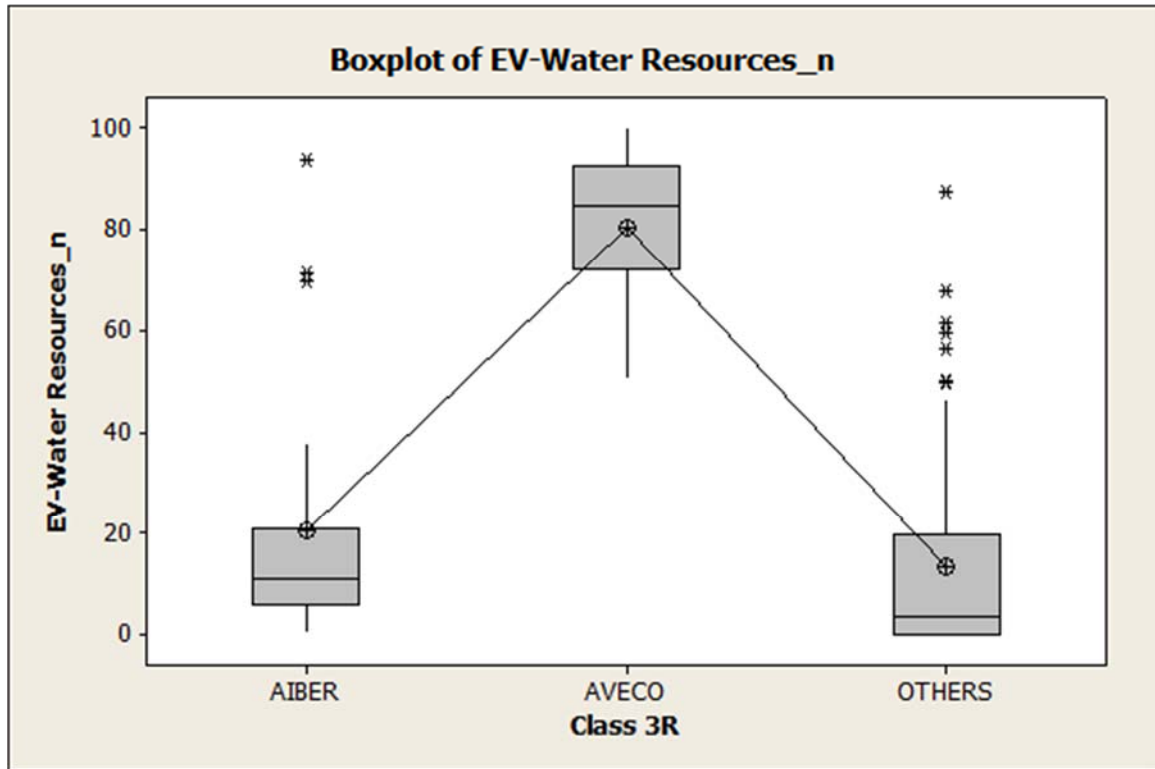
S = 19.73 R-Sq = 65.02% R-Sq(adj) = 64.48%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	CI Lower	CI Upper
AIBER	21	20.39	26.00	(---*---)	
AVECO	27	80.29	14.88		(---*---)
OTHERS	84	13.26	19.30	(--*-)	

Pooled StDev = 19.73

Boxplot of EV-Water Resources_n



Obs: *Outliers* que se destacaram positivamente foram Chile, Portugal e Espanha no agrupamento AIBER

One-way ANOVA: EV-Agriculture_n versus Class 6R

Source	DF	SS	MS	F	P
Class 6R	7	12631	1804	3.81	0.001
Error	124	58700	473		
Total	131	71330			

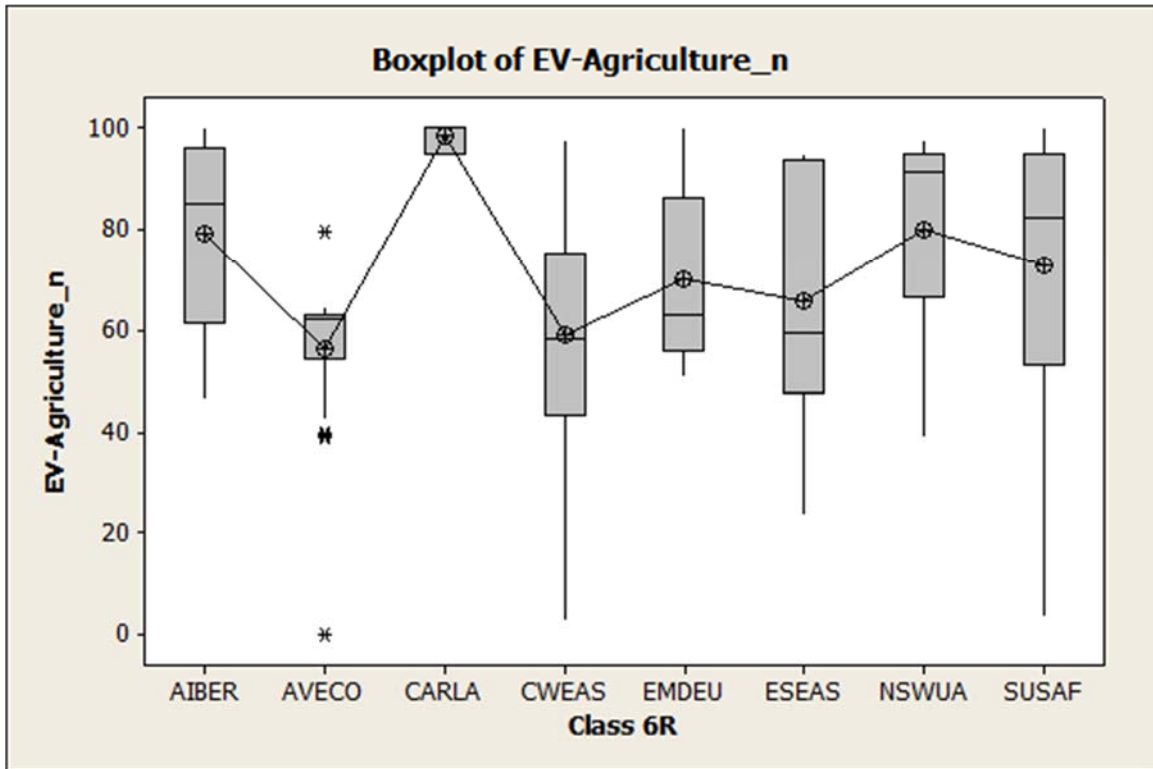
S = 21.76 R-Sq = 17.71% R-Sq(adj) = 13.06%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	CI Lower	CI Upper
AIBER	21	79.27	19.33	50.00	108.54
AVECO	27	56.11	14.68	26.84	85.38
CARLA	3	98.36	2.84	92.68	104.04
CWEAS	11	58.88	27.84	0.00	117.76
EMDEU	12	70.12	17.49	42.63	97.61
ESEAS	12	65.71	24.75	16.16	95.26
NSWUA	14	79.92	20.98	58.94	100.90
SUSAF	32	72.85	26.66	20.49	125.21

Pooled StDev = 21.76

Boxplot of EV-Agriculture_n



One-way ANOVA: EV-Agriculture_n versus Class 3R

Source	DF	SS	MS	F	P
Class 3R	2	7241	3621	7.29	0.001
Error	129	64089	497		
Total	131	71330			

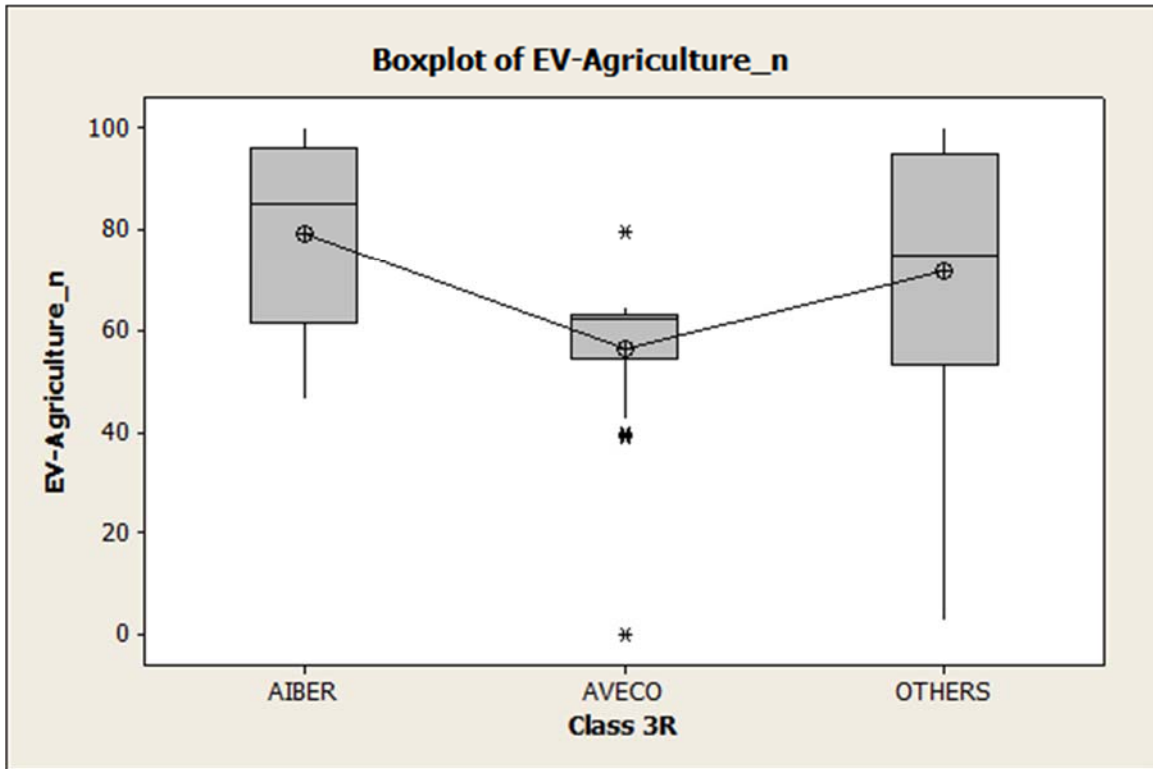
S = 22.29 R-Sq = 10.15% R-Sq(adj) = 8.76%

Level	N	Mean	StDev	95% CI
AIBER	21	79.27	19.33	(70.0, 88.5)
AVECO	27	56.11	14.68	(41.0, 71.2)
OTHERS	84	71.70	24.79	(67.0, 76.4)

Individual 95% CIs For Mean Based on Pooled StDev

Pooled StDev = 22.29

Boxplot of EV-Agriculture_n

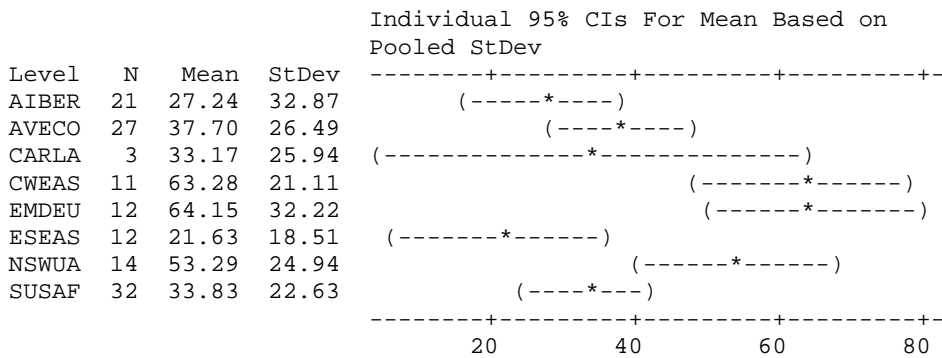


Obs: Os *outliers* no agrupamento AVECO em ordem crescente são: Israel, Japão, Coreia do Sul (Republic of Korea) e Nova Zelândia.

One-way ANOVA: EV-Forests_n versus Class 6R

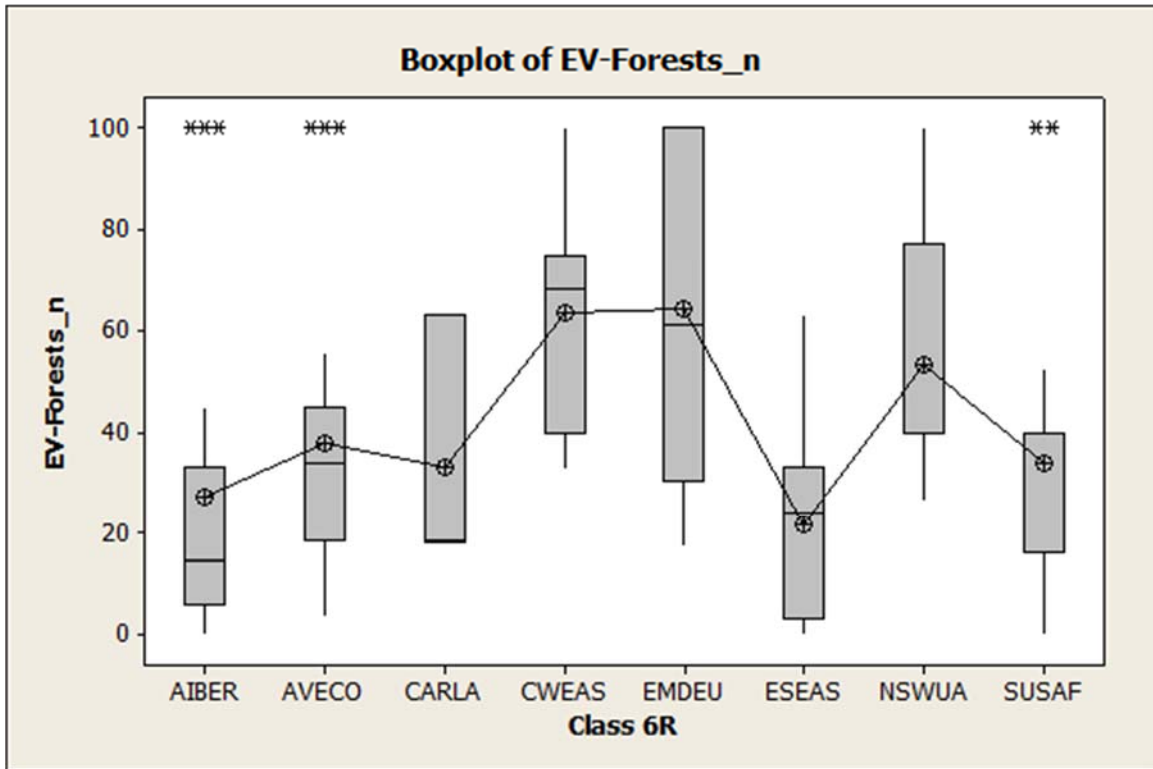
Source	DF	SS	MS	F	P
Class 6R	7	24394	3485	5.10	0.000
Error	124	84801	684		
Total	131	109195			

S = 26.15 R-Sq = 22.34% R-Sq(adj) = 17.96%



Pooled StDev = 26.15

Boxplot of EV-Forests_n

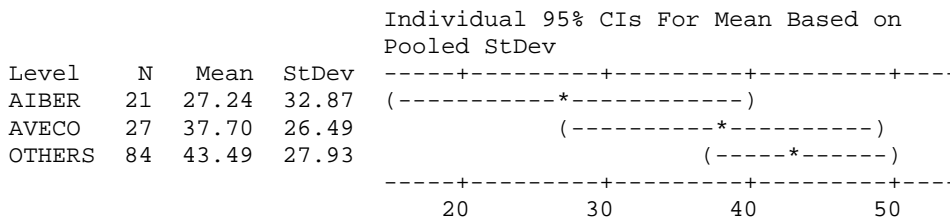


EV-Forests = 22 missing values

One-way ANOVA: EV-Forests_n versus Class 3R

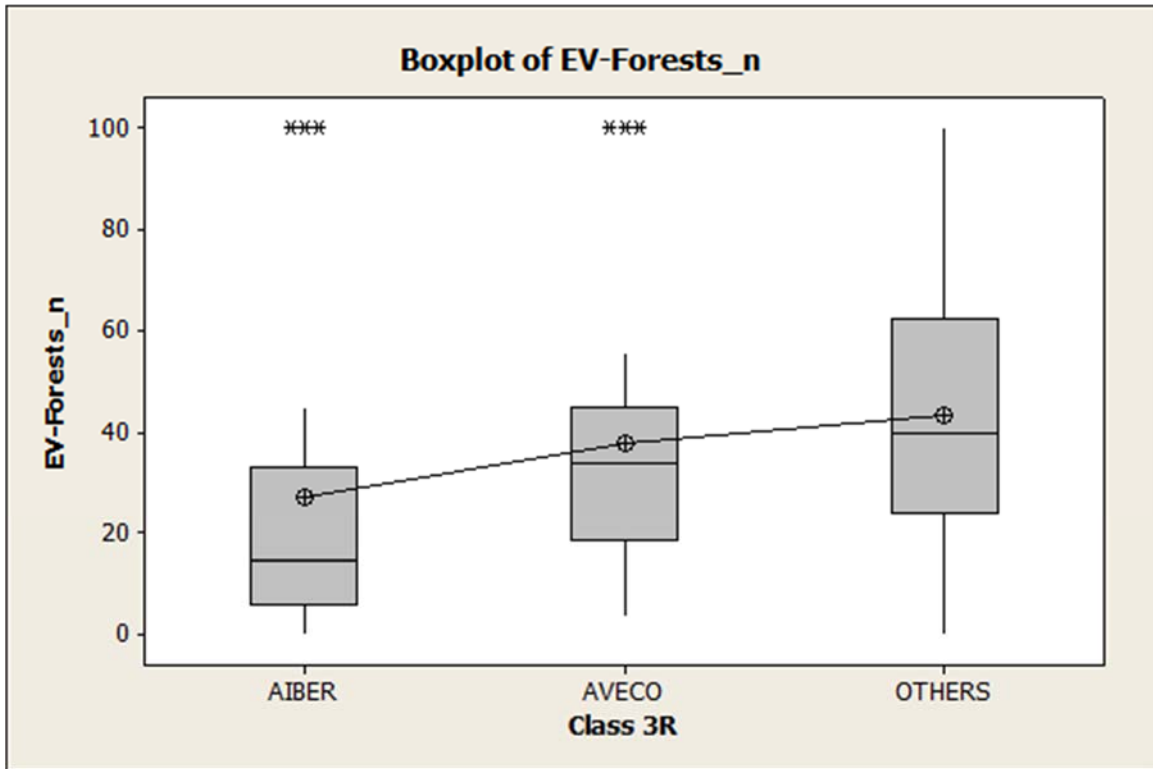
Source	DF	SS	MS	F	P
Class 3R	2	4577	2288	2.82	0.063
Error	129	104619	811		
Total	131	109195			

S = 28.48 R-Sq = 4.19% R-Sq(adj) = 2.71%



Pooled StDev = 28.48

Boxplot of EV-Forests_n



EV-Forests = 22 missing values

One-way ANOVA: EV-Fisheries_np versus Class 6R

Source	DF	SS	MS	F	P
Class 6R	7	6872	982	3.54	0.002
Error	124	34393	277		
Total	131	41265			

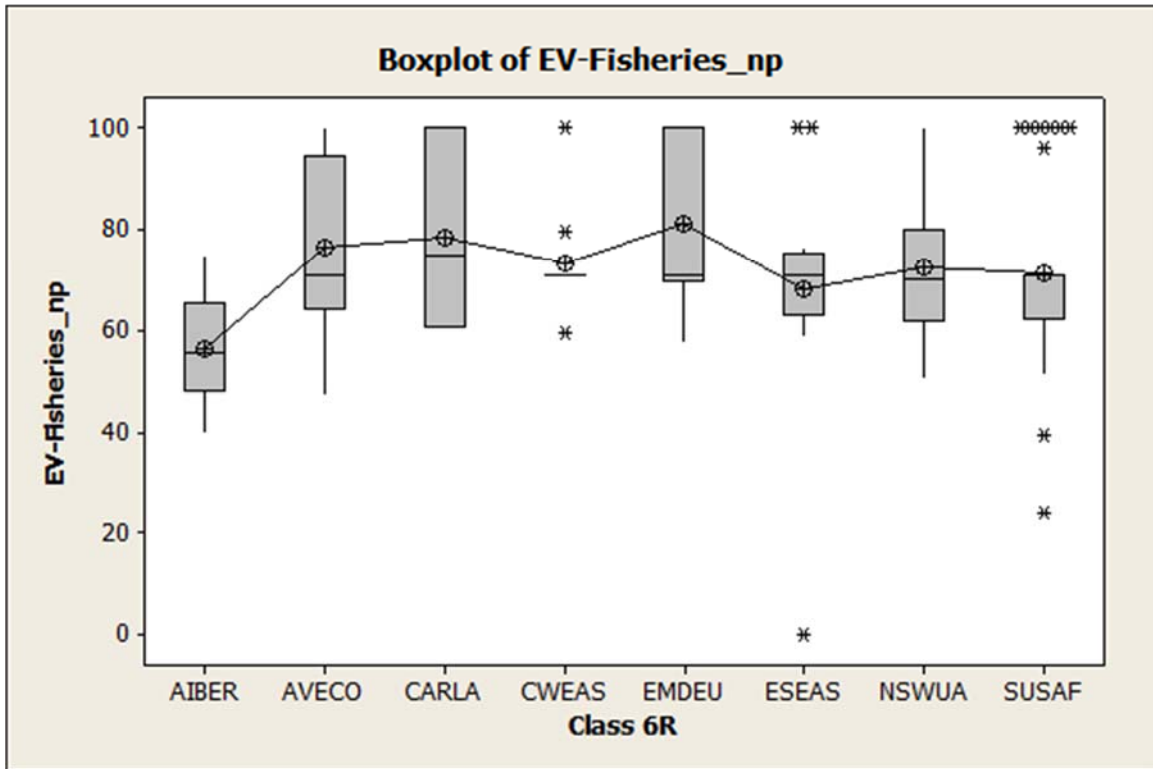
S = 16.65 R-Sq = 16.65% R-Sq(adj) = 11.95%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev
AIBER	21	56.39	10.71
AVECO	27	76.24	16.42
CARLA	3	78.45	19.95
CWEAS	11	73.39	9.93
EMDEU	12	81.30	16.93
ESEAS	12	68.01	25.08
NSWUA	14	72.57	14.82
SUSAF	32	71.32	18.30

Pooled StDev = 16.65

Boxplot of EV-Fisheries_np

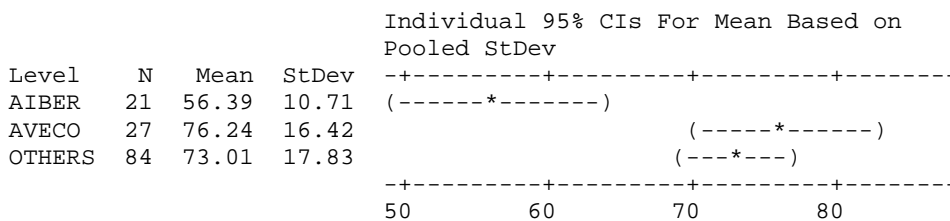


EV-Fisheries = 37 missing values

One-way ANOVA: EV-Fisheries_np versus Class 3R

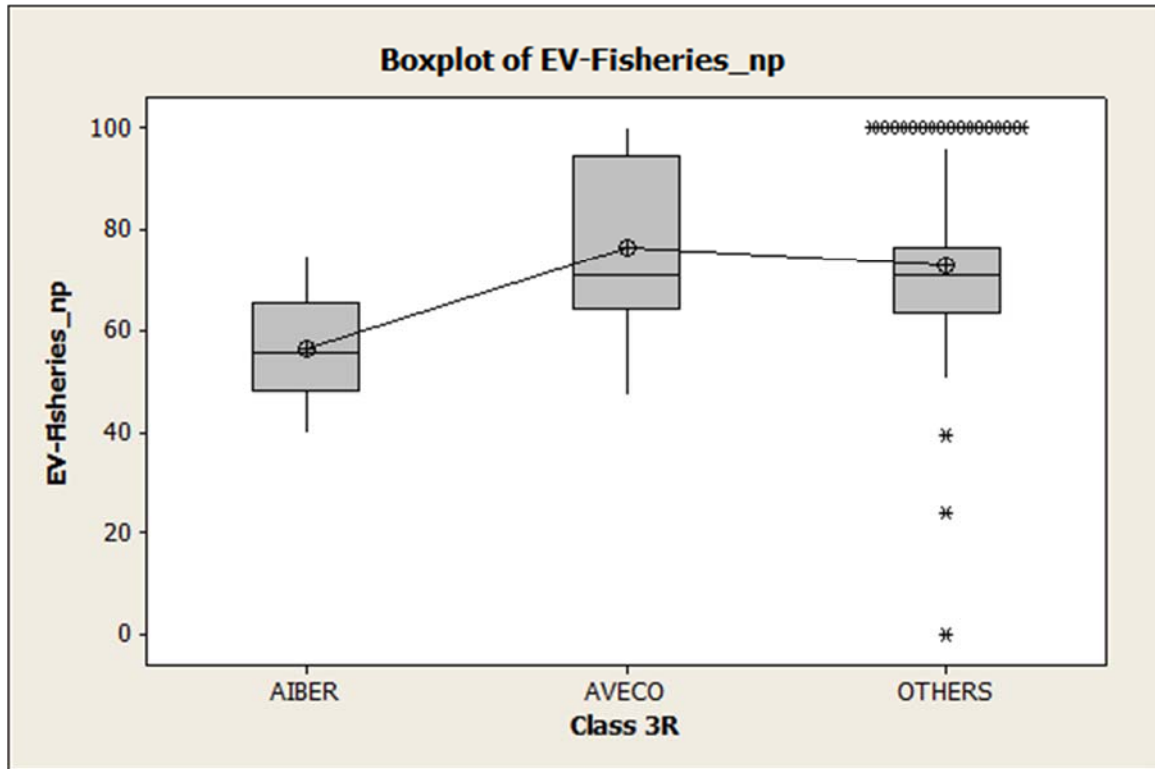
Source	DF	SS	MS	F	P
Class 3R	2	5563	2781	10.05	0.000
Error	129	35702	277		
Total	131	41265			

S = 16.64 R-Sq = 13.48% R-Sq(adj) = 12.14%



Pooled StDev = 16.64

Boxplot of EV-Fisheries_np



EV-Fisheries = 37 missing values

One-way ANOVA: Happy Planet Index_n versus Class 6R

Source	DF	SS	MS	F	P
Class 6R	7	28446	4064	15.53	0.000
Error	124	32448	262		
Total	131	60894			

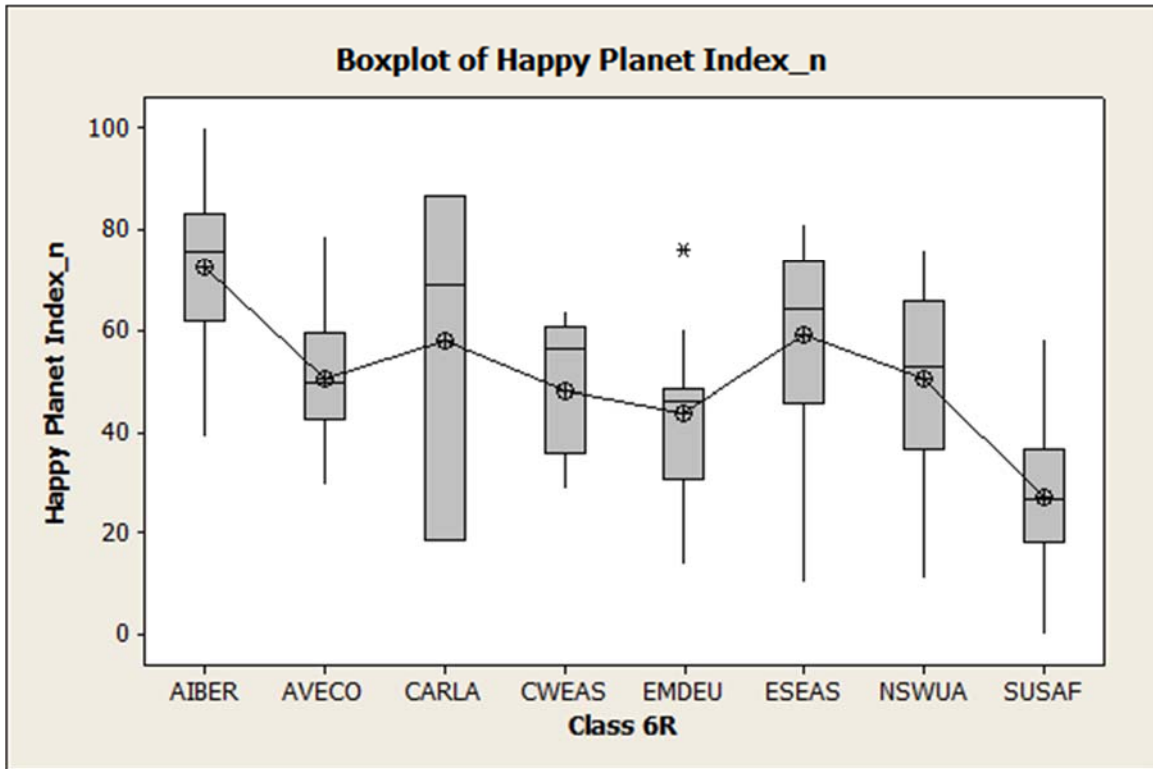
S = 16.18 R-Sq = 46.71% R-Sq(adj) = 43.71%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev
AIBER	21	72.36	16.25
AVECO	27	50.58	12.84
CARLA	3	58.07	35.38
CWEAS	11	48.13	13.54
EMDEU	12	43.56	16.02
ESEAS	12	59.06	20.02
NSWUA	14	50.22	19.91
SUSAF	32	27.27	14.09

Pooled StDev = 16.18

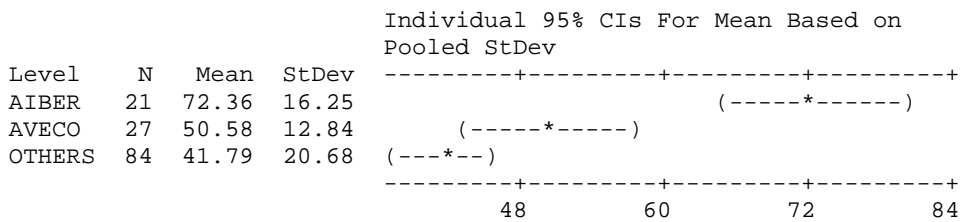
Boxplot of Happy Planet Index_n



One-way ANOVA: Happy Planet Index_n versus Class 3R

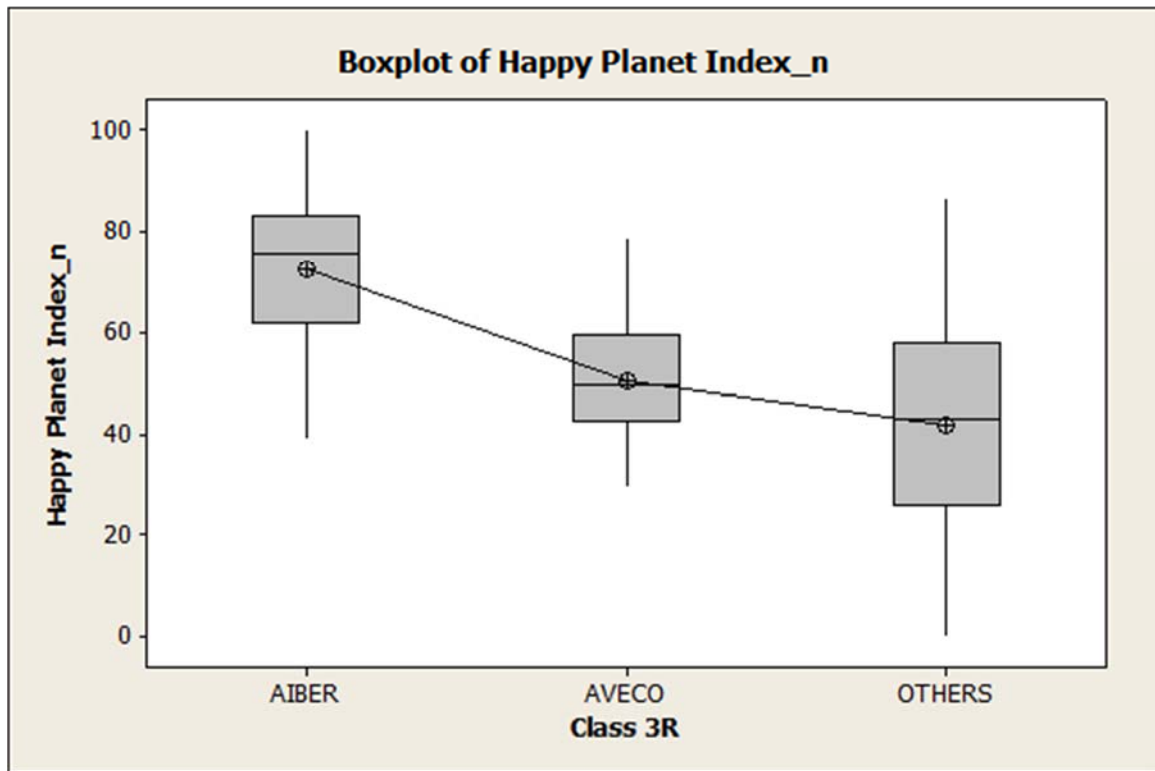
Source	DF	SS	MS	F	P
Class 3R	2	15846	7923	22.69	0.000
Error	129	45048	349		
Total	131	60894			

S = 18.69 R-Sq = 26.02% R-Sq(adj) = 24.87%



Pooled StDev = 18.69

Boxplot of Happy Planet Index_n

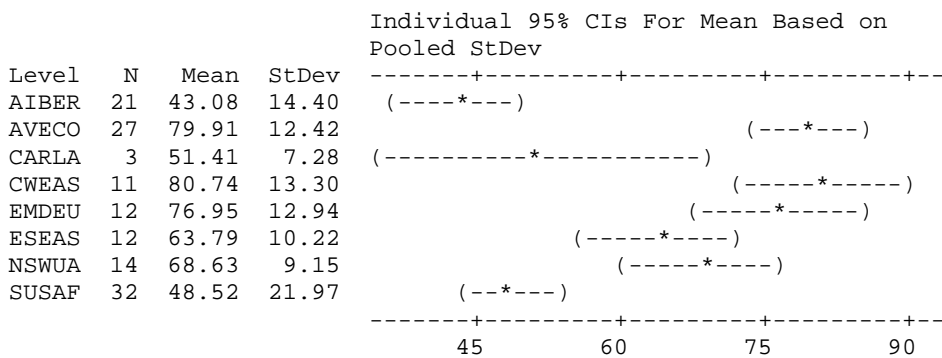


Obs: O outlier desapareceu e a conformação dos 3 agrupamentos permaneceram bem similares em relação ao gráfica anterior (class 6R).

One-way ANOVA: GINI Index_np versus Class 6R

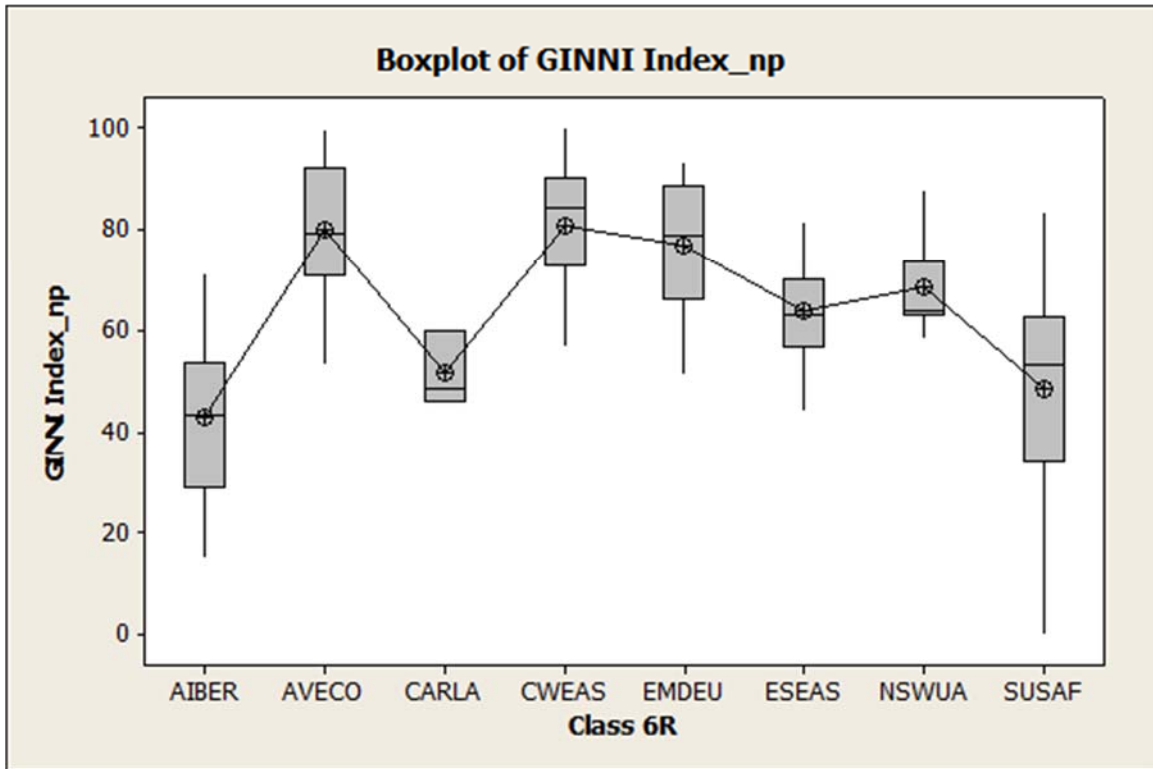
Source	DF	SS	MS	F	P
Class 6R	7	29417	4202	17.92	0.000
Error	124	29075	234		
Total	131	58492			

S = 15.31 R-Sq = 50.29% R-Sq(adj) = 47.49%



Pooled StDev = 15.31

Boxplot of GINNI Index_np



Obs: GINNI Index = 13 missing values

One-way ANOVA: GINNI Index_np versus Class 3R

Source	DF	SS	MS	F	P
Class 3R	2	16078	8039	24.45	0.000
Error	129	42414	329		
Total	131	58492			

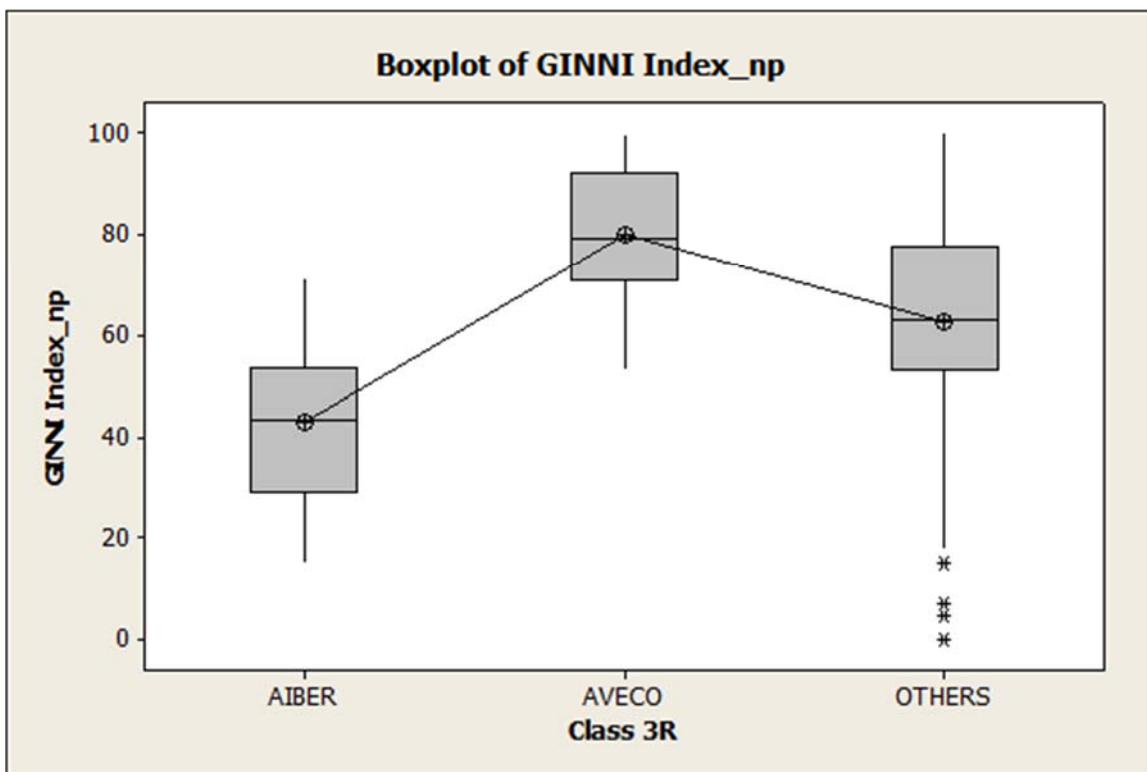
S = 18.13 R-Sq = 27.49% R-Sq(adj) = 26.36%

Level	N	Mean	StDev
AIBER	21	43.08	14.40
AVECO	27	79.91	12.42
OTHERS	84	62.44	20.32

Individual 95% CIs For Mean Based on Pooled StDev

Pooled StDev = 18.13

Boxplot of GINNI Index_np

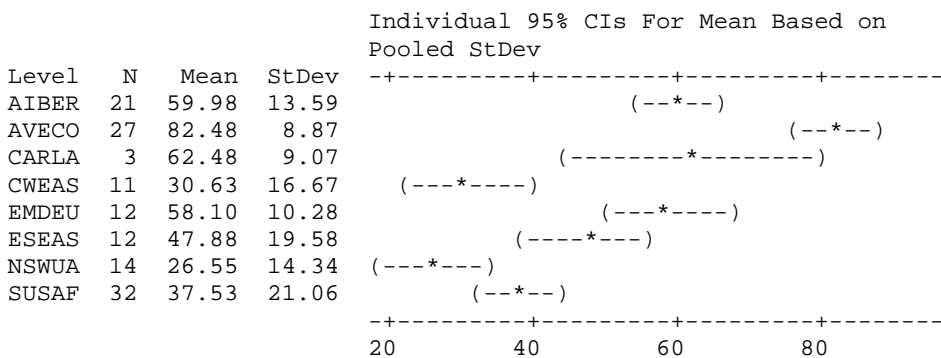


Obs: GINNI Index = 13 missing values

One-way ANOVA: Democracy Index-2014_n versus Class 6R

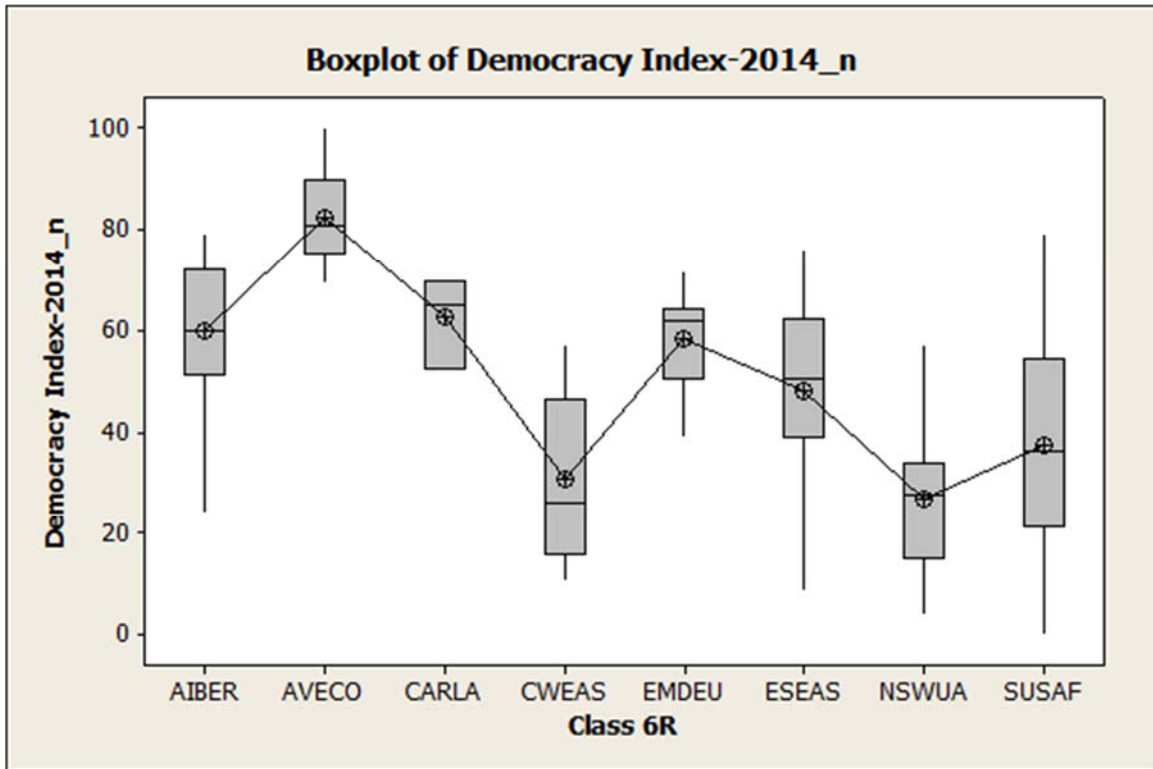
Source	DF	SS	MS	F	P
Class 6R	7	48182	6883	27.99	0.000
Error	124	30490	246		
Total	131	78672			

S = 15.68 R-Sq = 61.24% R-Sq(adj) = 59.06%



Pooled StDev = 15.68

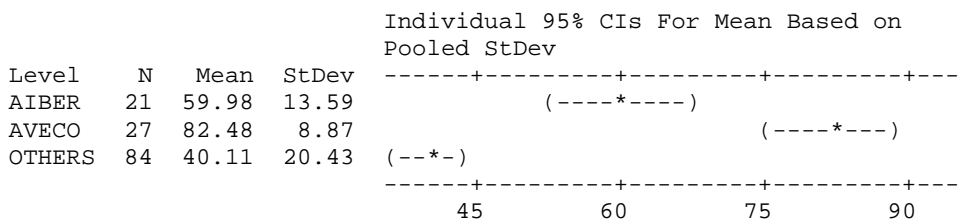
Boxplot of Democracy Index-2014_n



One-way ANOVA: Democracy Index-2014_n versus Class 3R

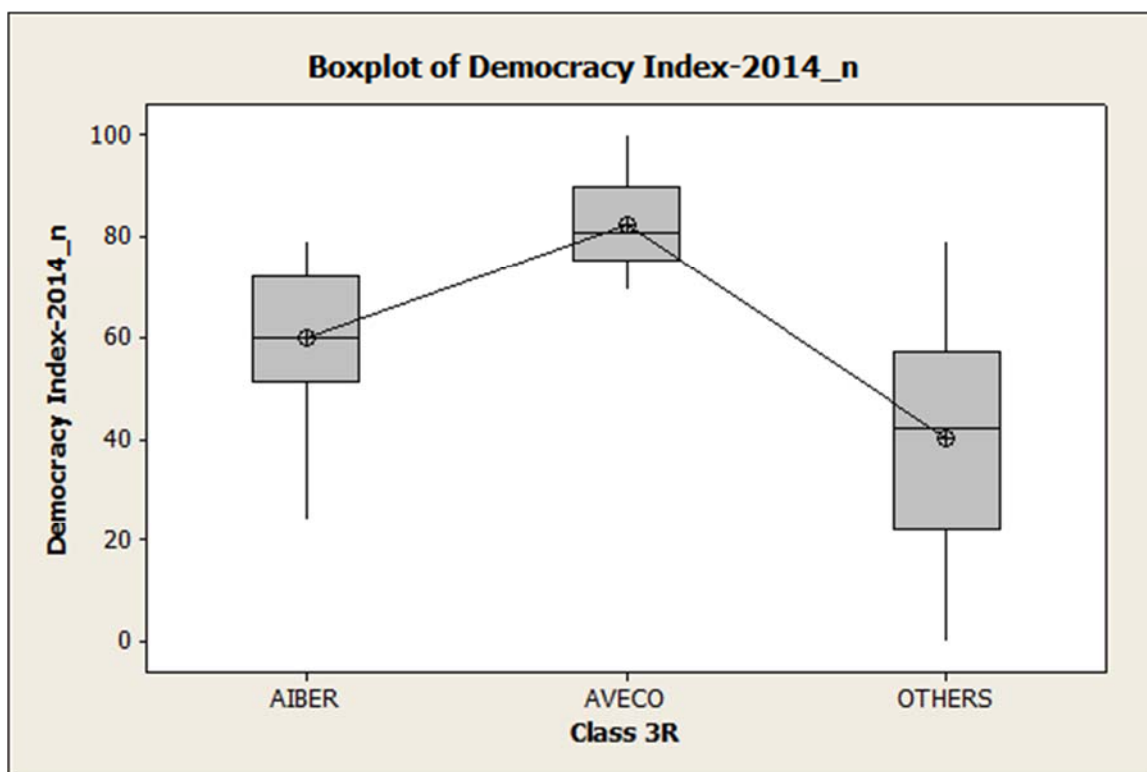
Source	DF	SS	MS	F	P
Class 3R	2	38300	19150	61.19	0.000
Error	129	40372	313		
Total	131	78672			

S = 17.69 R-Sq = 48.68% R-Sq(adj) = 47.89%



Pooled StDev = 17.69

Boxplot of Democracy Index-2014_n



5.4. Análise do Grau de discriminação e das Médias

Nesta etapa, todos os resultados da função One-Way ANOVA, foram ordenados considerando do maior F (grau de discriminação) para o menor. Em seguida, foram separados nas categorias: Indicador Sintético e Indicador Analítico. Seleccionados 7 indicadores mais relevantes (maior F) de cada categoria e elaborado as tabelas e gráficos a seguir:

Quadro: Médias e F dos Indicadores Sintéticos

Regiões	evwater	oppor	gover	spi	episc	fowell	democ
AIBER	20.39	55.28	42.48	64.41	52.70	63.48	59.98
AVECO	80.29	79.94	83.25	90.30	81.39	83.29	82.48
OTHERS	13.26	32.45	33.36	42.63	37.87	42.32	40.11
F	119.91	104.39	97.02	83.97	75.83	64.83	61.19

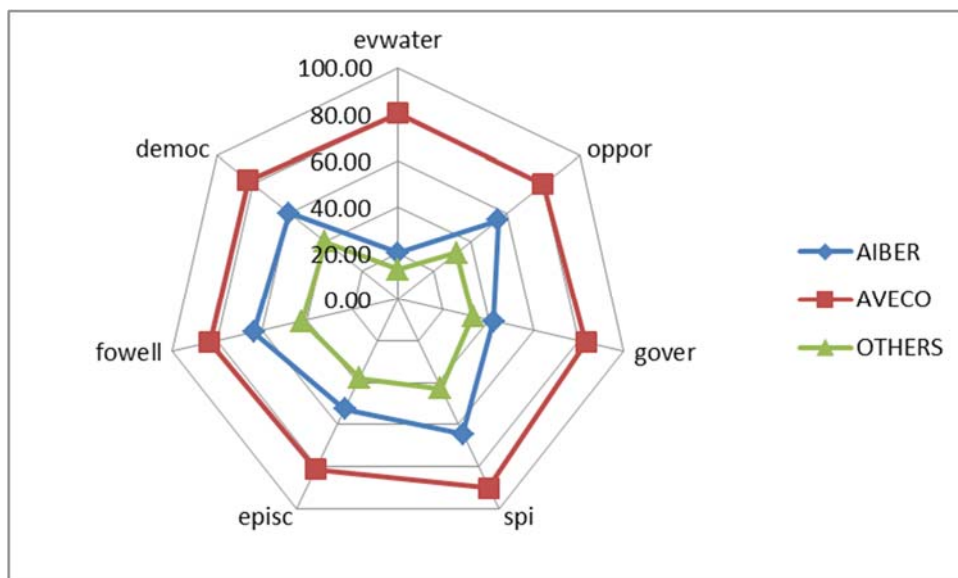


Gráfico com as Médias e F dos Indicadores Sintéticos

LEGENDA:

evwater: EV-Water Resources
 oppor: Opportunity
 gover: Governança
 spi: Social Progress Index
 episc: EPI Score
 fowell: Foundations of Wellbeing
 democ: Democracy Index-2014

Quadro: Médias e F dos Indicadores Analíticos

Regiões	tolhomo	freedom	yr3rds	wtreres	safenet	wyrscs	contra
AIBER	57.14	73.74	25.74	26.57	81.30	63.61	86.69
AVECO	64.51	85.82	49.63	70.33	89.43	86.80	88.99
OTHERS	16.64	56.65	18.05	56.16	65.13	49.92	62.11
F	103.30	68.49	37.45	33.17	28.77	26.41	21.15

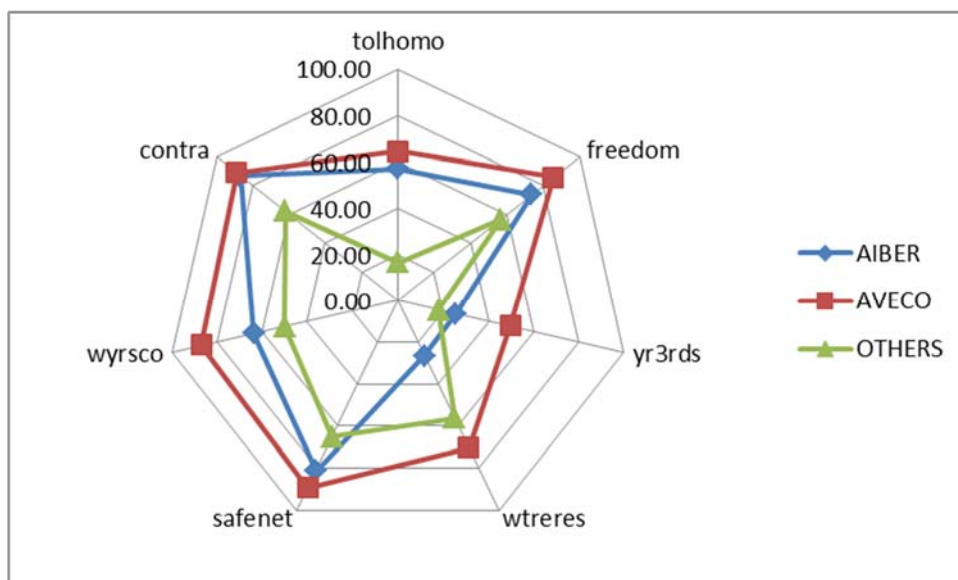


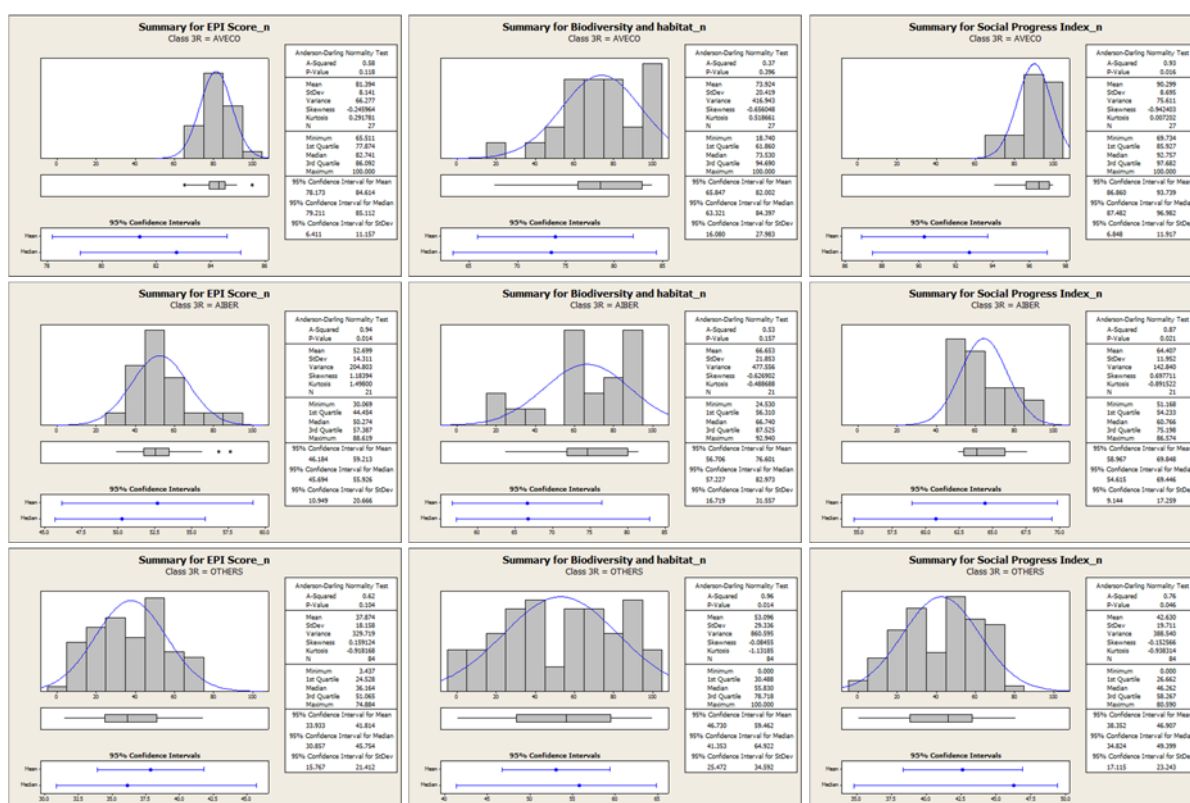
Gráfico com as Médias e F dos Indicadores Analíticos

LEGENDA:

tolhomo: Tolerance for homosexuals
 freedom: Personal Freedom and Choice
 yr3rds: Years of tertiary schooling
 wtreres: Women treated with respect
 safenet : Community safety net
 wyrSCO: Women`s years in school
 contra: Satisfied contraception (% of women)

5.5. Análise das novas variáveis (normalizadas e positivadas)

Os gráficos abaixo são resultados de análise descritiva para 3 (três) variáveis importantes referentes ao tema Biodiversidade – Capital Natural e Social. A análise foi feita por agrupamento dos países nas 3 (três) regiões objeto deste estudo, AIBER, AVECO e OTHERS. Foram dispostos verticalmente afim de tentar evidenciar tendências.

**5.6. CONSIDERAÇÕES**

O objetivo desta sessão foi analisar comparativamente as variáveis deste estudo quando agrupadas em 3 grandes blocos, a saber, AIBER com os países da região Ibero-

Americana, AVECO, os países de economias avançadas e OTHERS, os demais países constantes na base de dados, com ênfase no primeiro bloco.

Cuba foi reclassificada e passou a figurar como um país do bloco AIBER.

Na primeira fase foi utilizado o modelo ANOVA (one-way), para efetuar as comparações. Esta análise de variância testa a hipótese de que as médias de duas ou mais populações são iguais.

Das variáveis que apresentaram relevância, destacamos a “Necessidades Humanas Básicas”, com F de 41.59 e o país Letônia (Letônia) figura como *outlier* no agrupamento de países da região AVECO, para esta variável. No indicador “Fundamentos de Bem-Estar”, Cuba figura como *outlier* no agrupamento de países da região AIBER e F de 64.83 para esta variável.

A Letônia (Latvia), aparece como outlier na avaliação da variável “Saúde e Bem-Estar”, denotando resultado bem abaixo dos de países do agrupamento AVECO. E no grupo OTHERS, os dois outliers são Rússia e Cazaquistão (Kazakhstan). Com F relativamente baixo, de 32.23. Para o indicador “Liberdade Pessoal e de Escolha”, o F é de 68.49 e no agrupamento de países AIBER, Cuba aparece como outlier em relação aos demais. Os outliers do grupo OTHERS são respectivamente Paquistão, Sudão e Mauritània.

Para a variável que mede a “Emissão de Gases de Efeito Estufa”, a República Centro-Africana (Central African Republic), aparece neste estudo como o maior emissor de gases de efeito estufa. Segundo o relatório da ONU, sobre os dados de 2010, a principal causa deste resultado estava ligado a Desmatamento e Exploração de madeira para comércio (89.46%), seguido Agricultura (5.26%), entre outros. O F de 2.60, um dos mais baixos encontrados neste trabalho. Para o indicador de “Escravidão nos tempos atuais”, vemos no agrupamento AVECO, 3 *outliers* que destoam do grupo, são eles: Eslovênia, Eslováquia e República Tcheca. A Mauritània lidera o ranking nesta categoria, com cerca de 155,600 indivíduos (4% da pop.), em situação de escravidão.

fonte:<http://www.globallslaveryindex.org/country/mauritania/>

Com F de 21.15, o indicador que mede o “% de Mulheres Satisfeitas com os Métodos Contraceptivos” oferecidos em seus países, identificamos no agrupamento AIBER, a Guatemala como *outlier*, e no AVECO figuram Japão e Islândia respectivamente.

Para a variável que mede a percepção das “Mulheres que são tratadas com respeito” em seus países, listamos em AIBER, os 3 países com os menores valores, são eles: Perú com 24, República Dominicana com 18 e Colômbia com 17. O Brasil com 28, situa-se próximo destes. O F é de 33.17. Na análise de países para o índice de “Discriminação e Violência”,

apresentando F de 27.61, Israel aparece como outlier destoando significativamente de seu grupo (AVECO).

No indicador de “Segurança em Comunidade”, com F de 28.77, os 2 outliers que aparecem no agrupamento AVECO, são: Grécia e Coreia do Sul (Republic of Korea). A variável que representa a quantidade de “Anos das Mulheres na Escola”, a Islândia figura como outlier no agrupamento AVECO, o F desta análise é de 26.41

O “Índice de Desenvolvimento Humano-IDH”, com F de 34.45, destacou negativamente a Coreia do Sul (Republic of Korea), Letônia (Latvia), Holanda e Suíça como *outliers* na análise. No entanto, as considerações sobre estes resultados (*outliers*), não deverão ser levadas em conta, neste momento, haja vista que foi necessário ajuste nos valores de 3 países. Para os próximos trabalhos, estes casos passarão a refletir a as correções efetuadas.

O indicador EPI – Índice de Performance Ambiental, apresentou F de 75,83 e se destacaram positivamente como *outliers* os países Portugal e Espanha, os únicos pertencentes ao continente europeu dos países Ibero-Americanos e Suíça no AVECO. E negativamente como outlier a Coreia do Sul (Republic of Korea).

A variável *EV-Water Resources*, que reflete a preocupação quanto ao tratamento de águas residuais antes que sejam devolvidas ao meio-ambiente, apresentou um F de 119.91, o mais alto deste estudo em relação às outras variáveis. Como *outliers* que se destacaram positivamente temos os países Chile, Portugal e Espanha no agrupamento AIBER. Por sua vez, o indicador *EV-Agriculture*, que retrata os países que recebem subsídios para sua produção, despontou como *outliers* no agrupamento AVECO em ordem crescente são: Israel, Japão, Coreia do Sul (Republic of Korea) e Nova Zelândia.

O Índice de Felicidade do Planeta – HPI, resultou em um F de 22.69 e foi o único indicador em que AIBER performou melhor que AVECO.

Para algumas variáveis não foi possível uma análise mais robusta devido a falta de dados, para citar algumas, temos *EV-Fisheries_np* (37 missing values), *EV-Forests* (22 missing values), *Years of tertiary schooling* (15 missing values) e *GINI Index* (13 missing values), entre outras.

Na sessão de Análise do grau de discriminação e das médias, onde foi utilizado os resultados da função *One-Way ANOVA*, foi possível visualizar graficamente de maneira sintetizada o que os boxplots já vinham indicando individualmente para cada variável.

Em linhas gerais, os resultados deste trabalho de comparações com agrupamentos por regiões, veio reforçar o desempenho superior dos países AVECO em relação ao AIBER e demais países objetos deste estudo.

6. AMOSTRAGEM

Os objetivos desta sessão são, realizar uma pesquisa por amostragem dos dados dimensionadores para as variáveis SPI-2014 – Índice de Progresso Social, IDH-2014 – Índice de Desenvolvimento Humano e EPI-2014 – Índice de Performance Ambiental, efetuar análises ANOVA para verificar variância entre AIBER e AVECO e testes de amostragens (two-sample) e gráficos boxplots.

O dados serão categorizados em 3 (três) principais regiões, os países Ibero-Americanos (AIBER), países de economias avançadas (AVECO) e os demais países constantes na base de dados deste estudo.

6.1 As variáveis

Este estudo trata de 4 (quatro) variáveis, sendo 3 (três) quantitativas – SPI-2014, IDH-2014 e EPI-2014, e 1 (um) categórica – Class 3R (região) descritas abaixo no Quadro 2:

Quadro 2. As 4 variáveis do eixo temático para este estudo

Variável	Significado	Tipo	Unidade de Medida
SPI-2014	O Índice de Progresso Social é composto por três dimensões do progresso social: Necessidades Humanas Básicas, Fundamentos de Bem-Estar e Oportunidades. fonte: www.socialprogressimperative.org	Variável quantitativa	Indicador nominal
IDH-2014	O Índice de Desenvolvimento Humano (IDH) mede o progresso de uma nação a partir de três dimensões: renda, saúde e educação.	Variável quantitativa	Indicador nominal
EPI-2014	Índice de Performance Ambiental, composto por: Saúde Ambiental (saúde humana e dos ecossistemas), Impactos na Saúde (medido pela mortalidade infantil – idade entre 1 e 5 anos), Qualidade do Ar (poluição do ar, partículas sólidas e seu impacto na saúde) e Nível de Água tratada e Sanitização adequada.	Variável quantitativa	Indicador nominal
Class 3R	Variável utilizada para agrupar os 132 países em 3 (três) regiões, AIBER (Ibero-Americana, AVECO (Economias Avançadas) e OTHERS (demais países)	Variável categórica	N/A

Fonte: O autor a partir dos dados da planilha estatística, do GPS p.16 e do site do Banco Mundial

6.2. PESQUISA POR AMOSTRAGEM

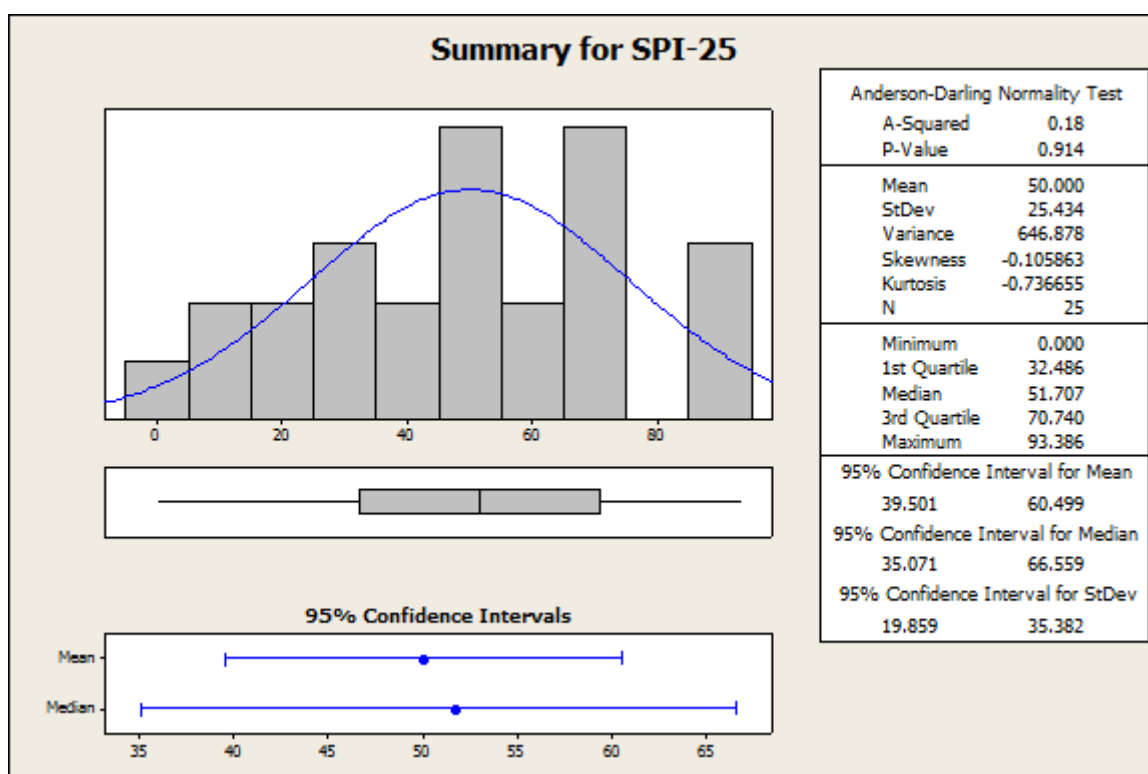
A pesquisa por amostragem foi feita em três amostras extraídas aleatoriamente da base de dados. Uma com 25 países, outra com 50 e por fim, uma amostra com 100 países. Estas amostras foram geradas a partir da funcionalidade “*Random Data*” do software Minitab. Para cada amostra extraída foram efetuadas análises exploratórias dos dados, análise de variância (ANOVA), testes de amostragens (two-sample) e gráficos boxplots.

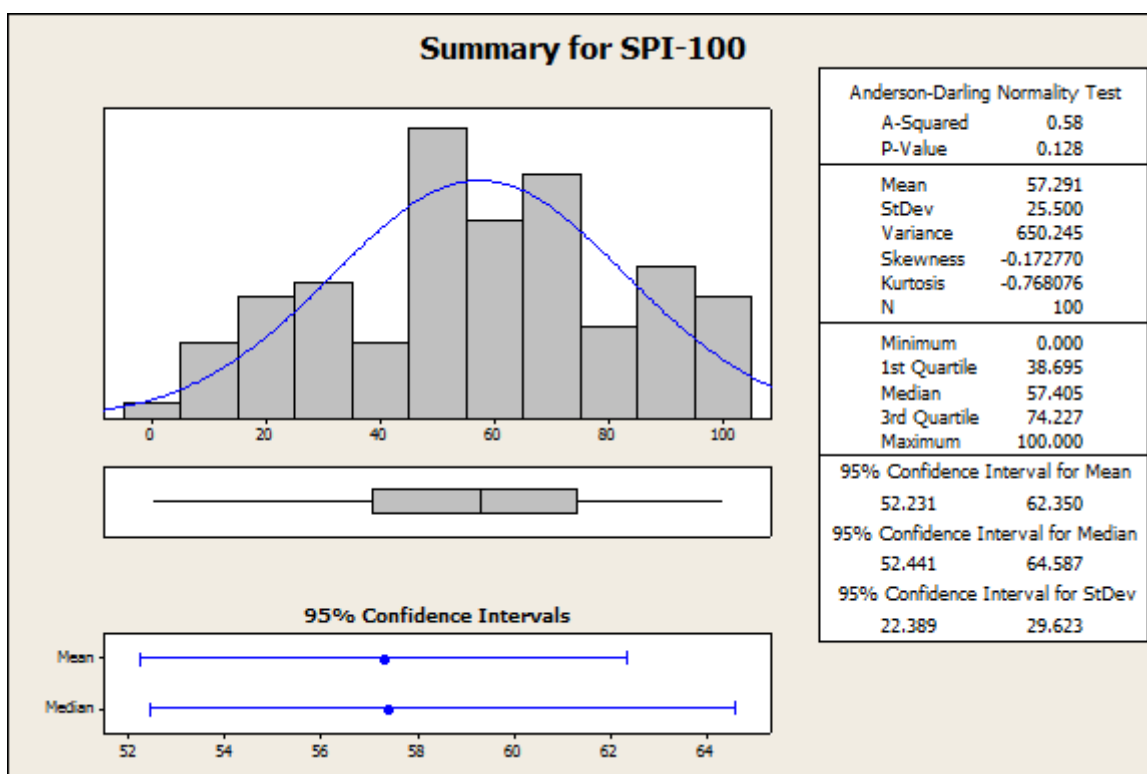
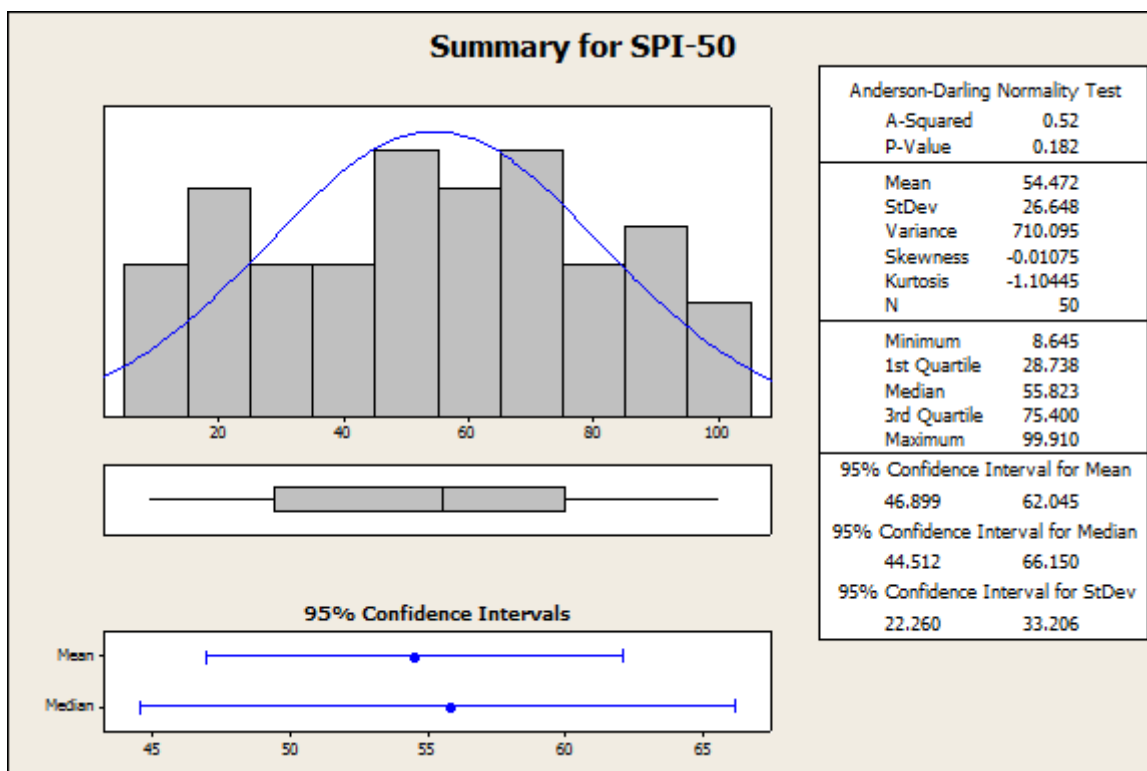
6.2.1 SPI-2014 – Índice de Progresso Social

Nas sessões que seguem, serão demonstrados os resultados da análise exploratória, análise de variância e teste de amostragem entre duas regiões para a variável SPI-2014.

6.2.1.1 Análise Exploratória

Abaixo os gráficos boxplots de análise exploratória com as amostragens de 25, 50 e 100 países, escolhidas aleatoriamente entre as 3 (três) regiões.





6.2.1.2 Análise de variância (ANOVA unstacked)

Os resultados para a ANOVA, considerando as amostras com 25, 50, 100 e 132 países, são demonstrados abaixo.

One-way ANOVA: SPI-25, SPI-50, SPI-100, SPI-2014_n

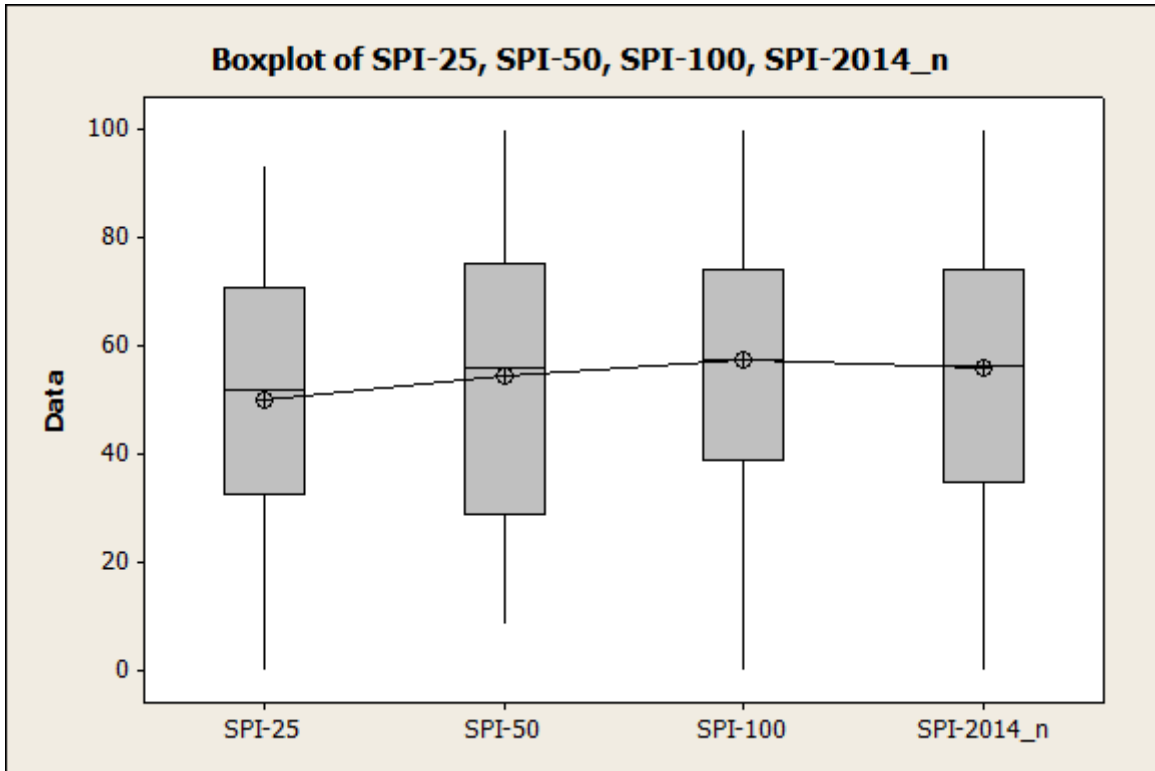
Source	DF	SS	MS	F	P
Factor	3	1141	380	0.58	0.631
Error	303	200027	660		
Total	306	201168			

S = 25.69 R-Sq = 0.57% R-Sq(adj) = 0.00%

Level	N	Mean	StDev	Individual 95% CIs For Mean Based on Pooled StDev
SPI-25	25	50.00	25.43	(-----*-----)
SPI-50	50	54.47	26.65	(-----*-----)
SPI-100	100	57.29	25.50	(-----*-----)
SPI-2014_n	132	55.84	25.52	(-----*-----)

-----+-----+-----+-----+-----
42.0 48.0 54.0 60.0

Pooled StDev = 25.69



Obs: No gráfico acima, a amostra com 100 países apresentou o menor intervalo, denotando representar melhor a totalidade da população, com 95% de confiança.

Comparação entre as regiões AIBER e AVECO (ANOVA unstacked)

One-way ANOVA: spi_AIBER, spi_AVECO

Source	DF	SS	MS	F	P
Factor	1	7919	7919	75.53	0.000
Error	46	4823	105		
Total	47	12742			

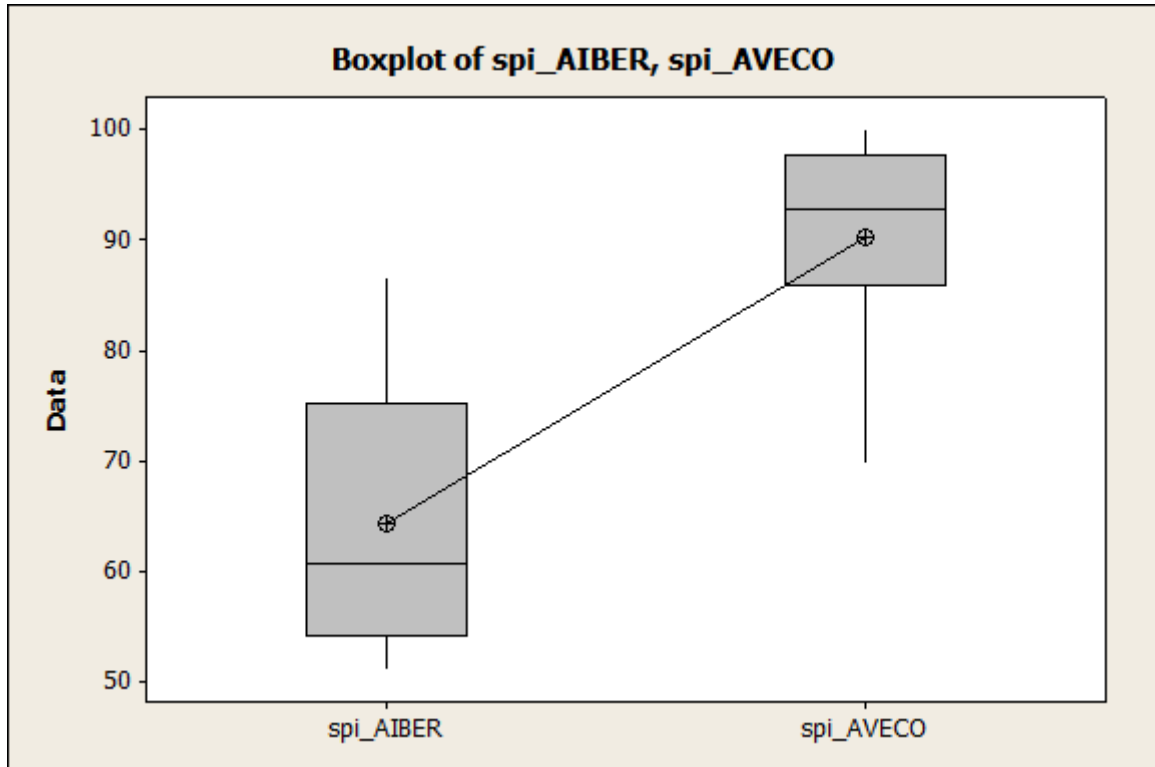
S = 10.24 R-Sq = 62.15% R-Sq(adj) = 61.33%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	+-----+-----+-----+-----
spi_AIBER	21	64.41	11.95	(---*---)
spi_AVECO	27	90.30	8.70	(---*---)

60 70 80 90

Pooled StDev = 10.24



6.2.1.3 Testes de amostragens (two-sample)

Teste de amostragens (T-Test).

Two-Sample T-Test and CI: spi_AIBER, spi_AVECO

Two-sample T for spi_AIBER vs spi_AVECO

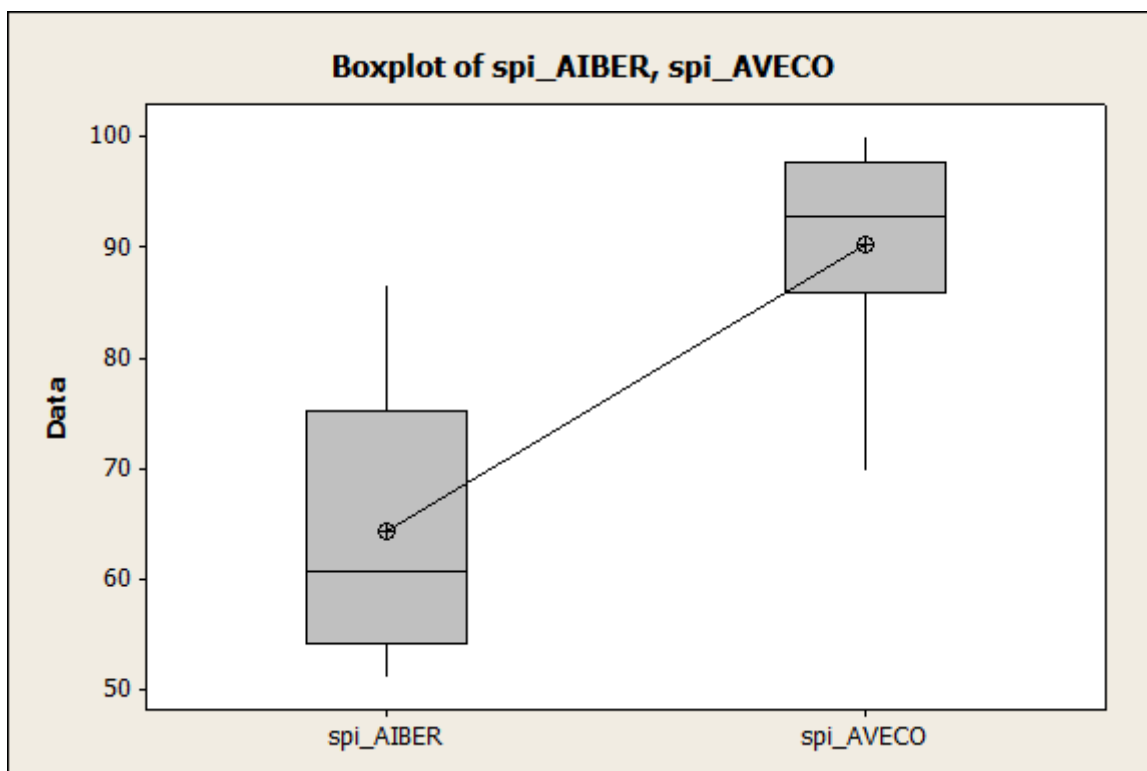
	N	Mean	StDev	SE Mean
spi_AIBER	21	64.4	12.0	2.6
spi_AVECO	27	90.30	8.70	1.7

Difference = mu (spi_AIBER) - mu (spi_AVECO)

Estimate for difference: -25.89

95% upper bound for difference: -20.66

T-Test of difference = 0 (vs <): T-Value = -8.36 P-Value = 0.000 DF = 35

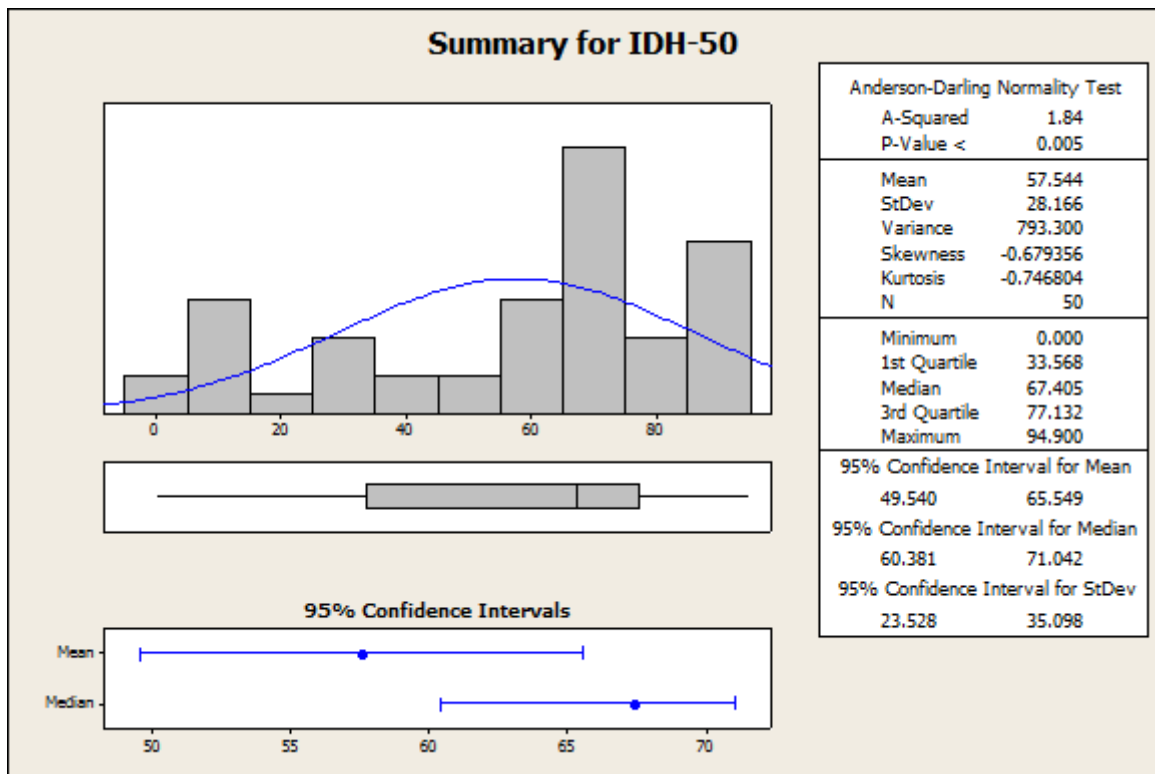
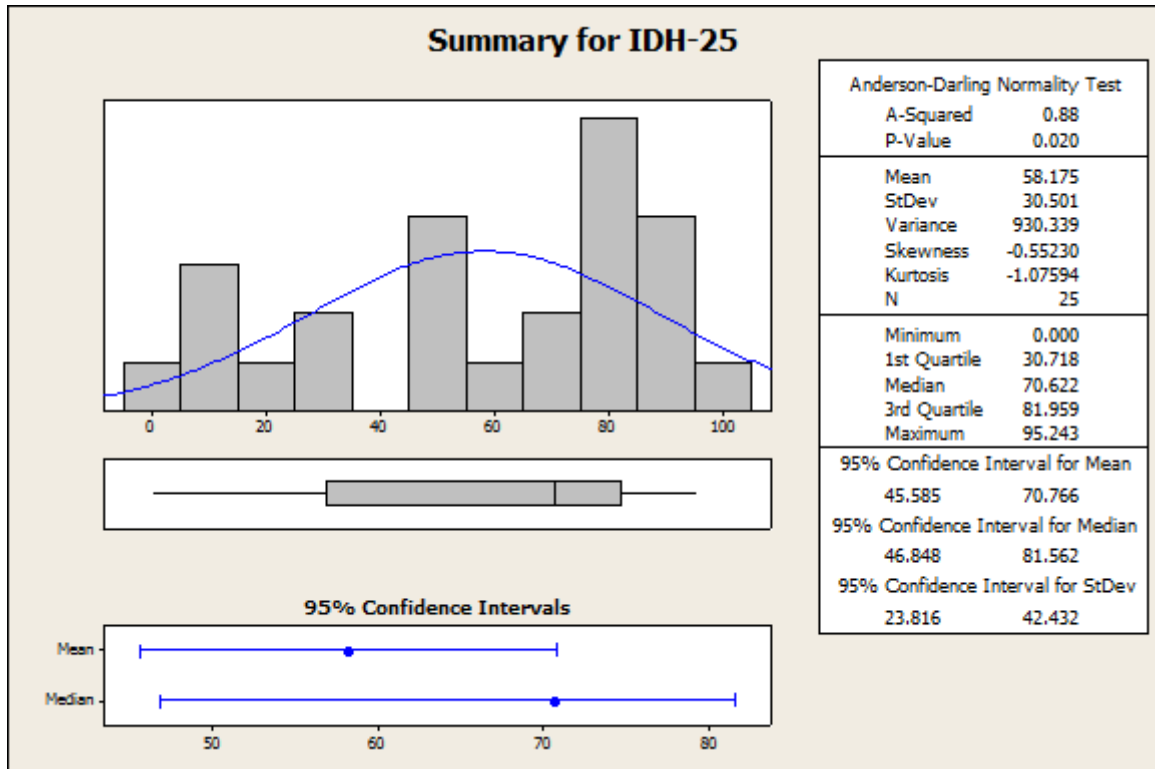


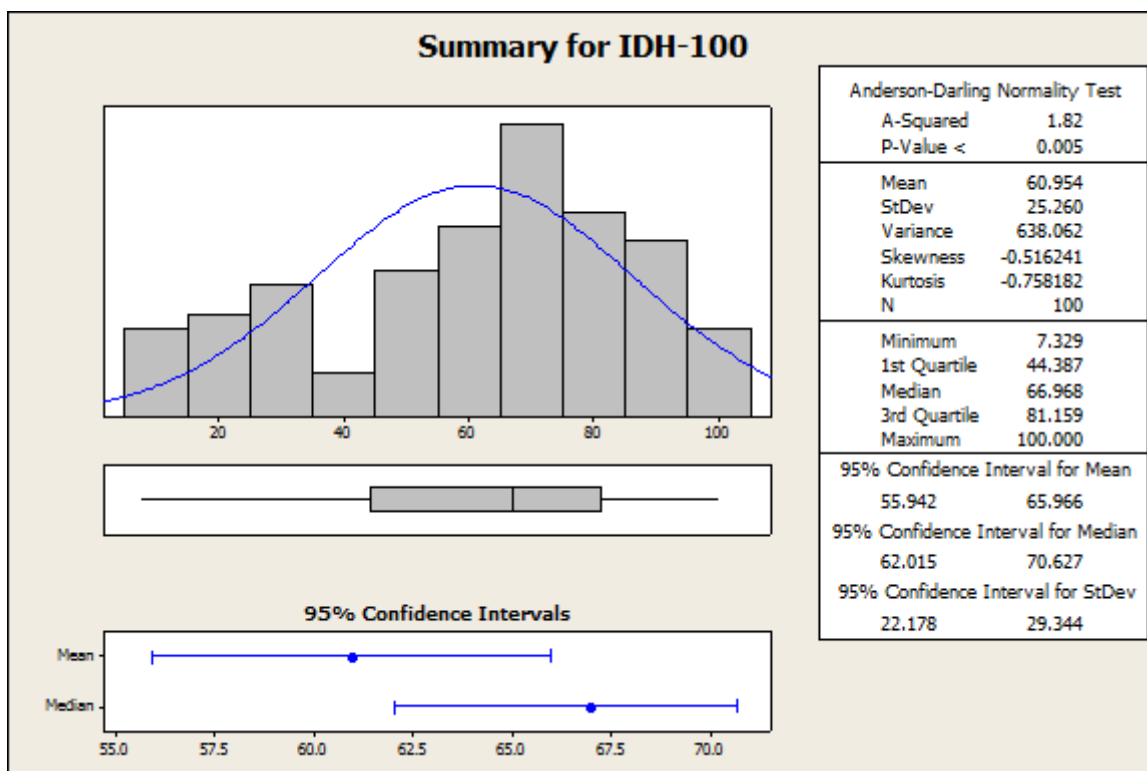
6.3 IDH-2014 – Índice de Desenvolvimento Humano

Nas sessões que seguem, serão demonstrados os resultados da análise exploratória, análise de variância e teste de amostragem entre duas regiões para a variável IDH-2014.

6.3.1.1 Análise Exploratória

Abaixo os gráficos boxplots de análise exploratória com as amostragens de 25, 50 e 100 países, escolhidas aleatoriamente entre as 3 (três) regiões.





6.3.1.2 Análise de variância (ANOVA unstacked)

Os resultados para a ANOVA, considerando as amostras com 25, 50, 100 e 132 países, são demonstrados abaixo.

One-way ANOVA: IDH-25, IDH-50, IDH-100, IDH-2014_n

Source	DF	SS	MS	F	P
Factor	3	446	149	0.21	0.891
Error	303	215937	713		
Total	306	216382			

S = 26.70 R-Sq = 0.21% R-Sq(adj) = 0.00%

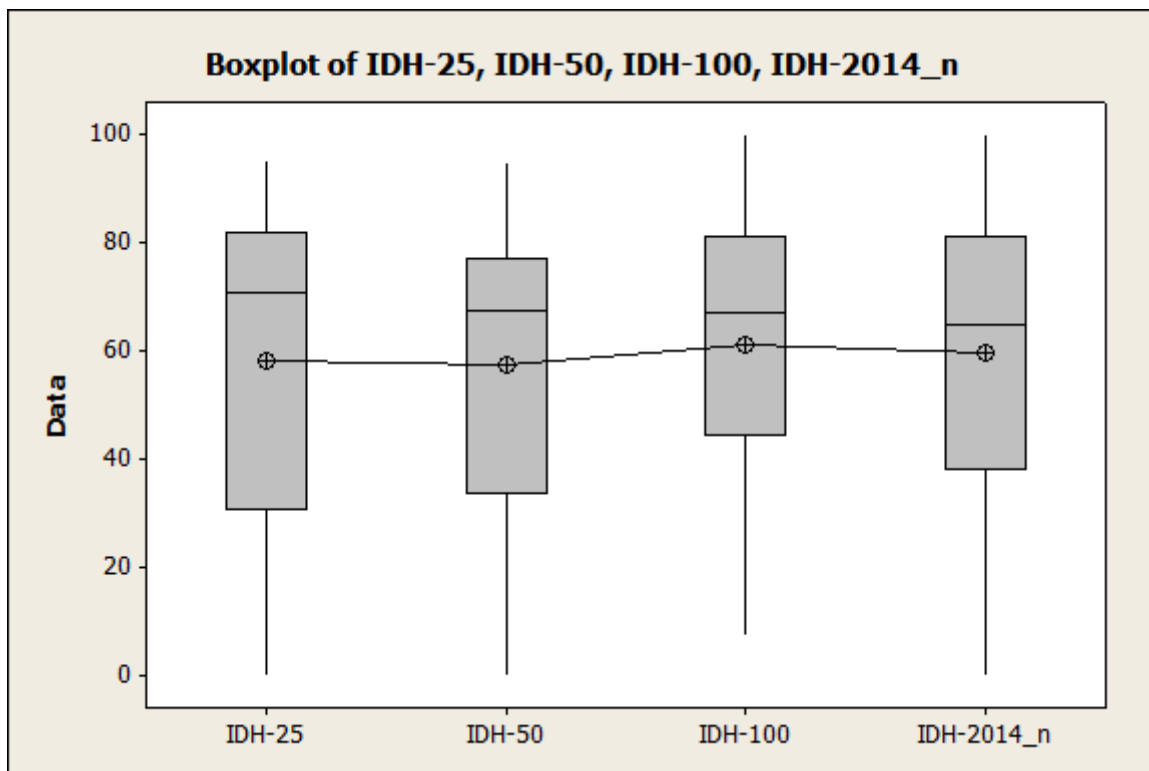
Level	N	Mean	StDev
IDH-25	25	58.18	30.50
IDH-50	50	57.54	28.17
IDH-100	100	60.95	25.26
IDH-2014_n	132	59.64	26.44

Individual 95% CIs For Mean Based on Pooled StDev

Level	Lower CI	Upper CI
IDH-25	22.178	29.344
IDH-50	22.178	29.344
IDH-100	22.178	29.344
IDH-2014_n	22.178	29.344

48.0 54.0 60.0 66.0

Pooled StDev = 26.70



Obs: No gráfico acima, a amostra com 100 países apresentou o menor intervalo, denotando representar melhor a totalidade da população, com 95% de confiança.

Comparação entre as regiões AIBER e AVECO (ANOVA unstacked)

One-way ANOVA: idh_AIBER, idh_AVECO

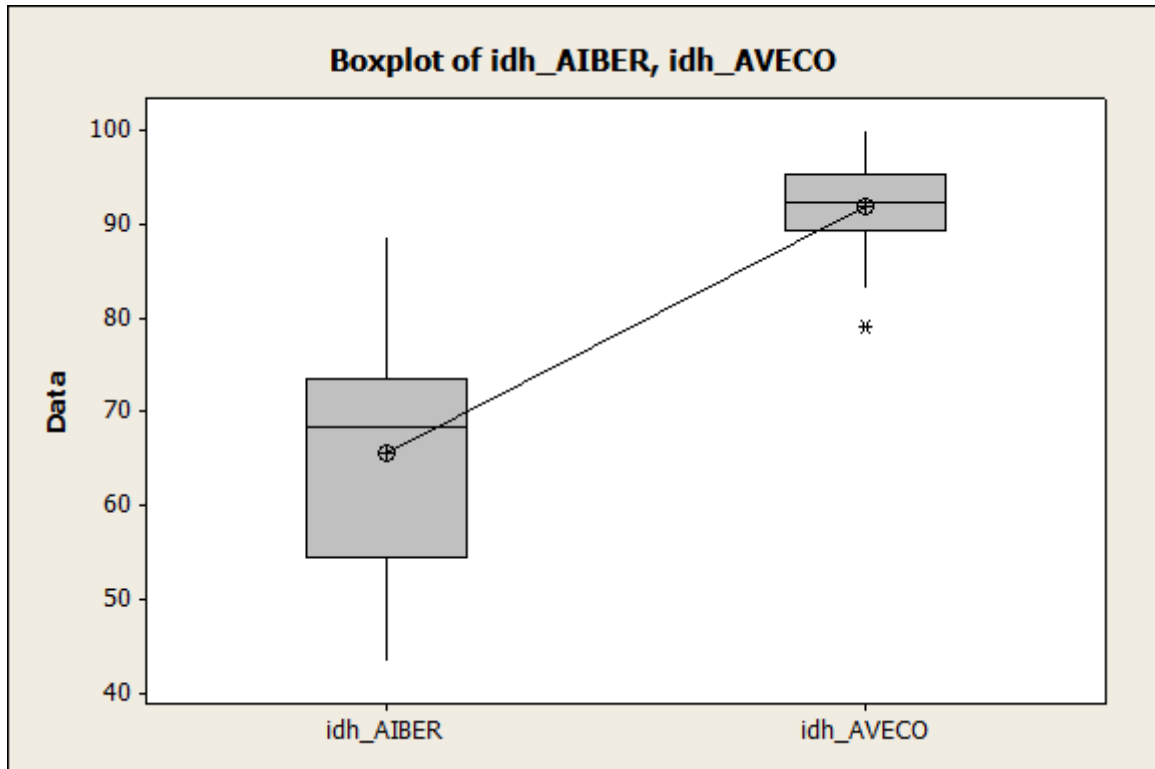
Source	DF	SS	MS	F	P
Factor	1	8122.1	8122.1	101.08	0.000
Error	46	3696.1	80.4		
Total	47	11818.2			

S = 8.964 R-Sq = 68.73% R-Sq(adj) = 68.05%

Level	N	Mean	StDev
idh_AIBER	21	65.671	12.452
idh_AVECO	27	91.892	4.784

Individual 95% CIs For Mean Based on Pooled StDev

Pooled StDev = 8.964



Obs: A Letônia (Latvia) figura como *outlier* no boxplot acima.

6.3.1.3 Testes de amostragens (two-sample)

Teste de amostragens (T-Test).

Two-Sample T-Test and CI: idh_AIBER, idh_AVECO

Two-sample T for idh_AIBER vs idh_AVECO

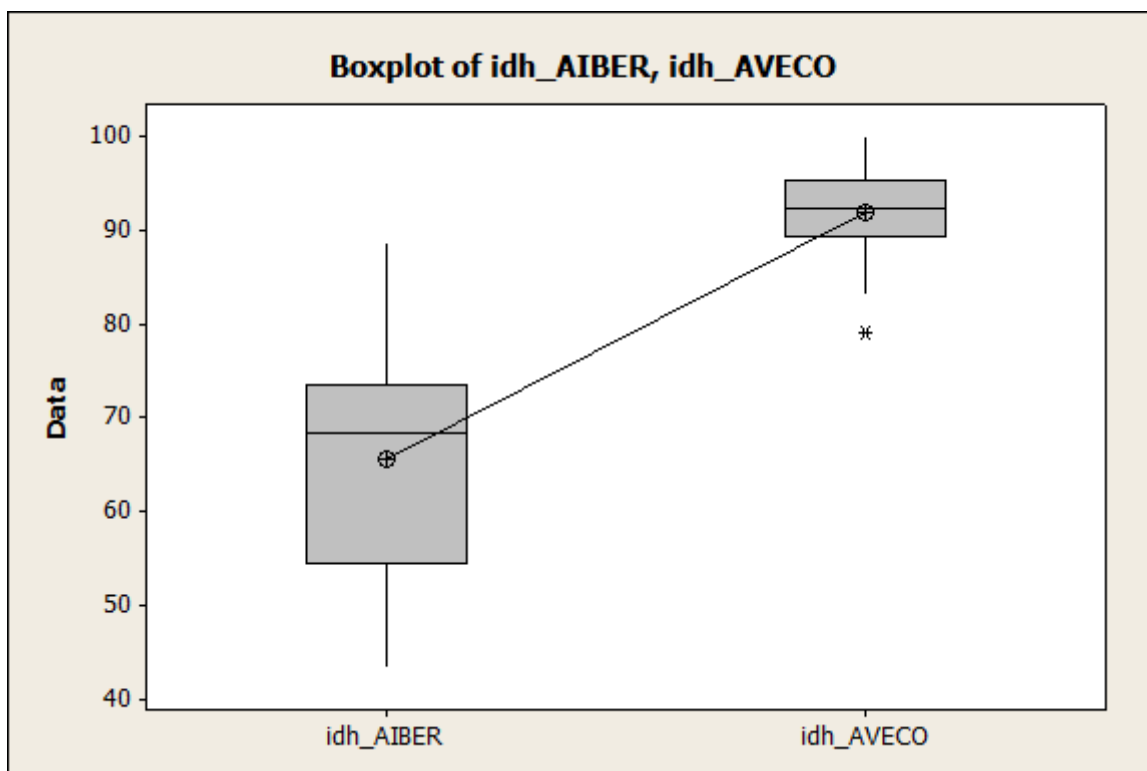
	N	Mean	StDev	SE Mean
idh_AIBER	21	65.7	12.5	2.7
idh_AVECO	27	91.89	4.78	0.92

Difference = mu (idh_AIBER) - mu (idh_AVECO)

Estimate for difference: -26.22

95% upper bound for difference: -21.31

T-Test of difference = 0 (vs <): T-Value = -9.14 P-Value = 0.000 DF = 24



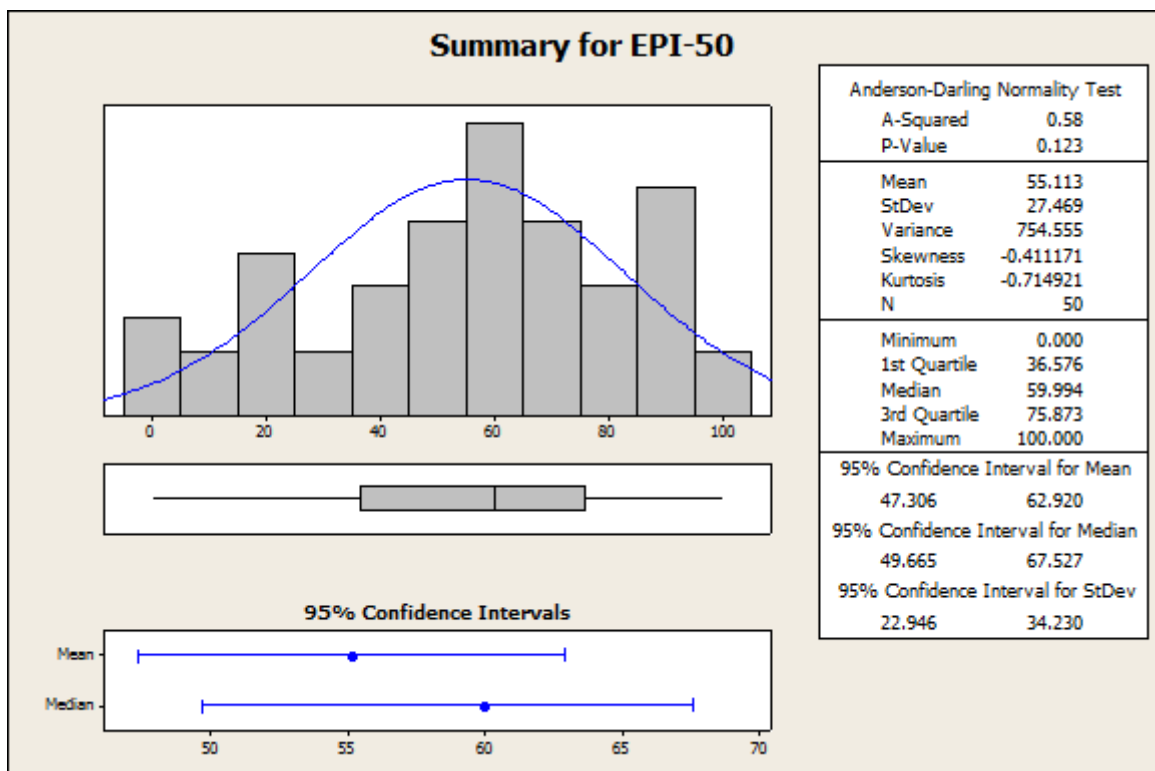
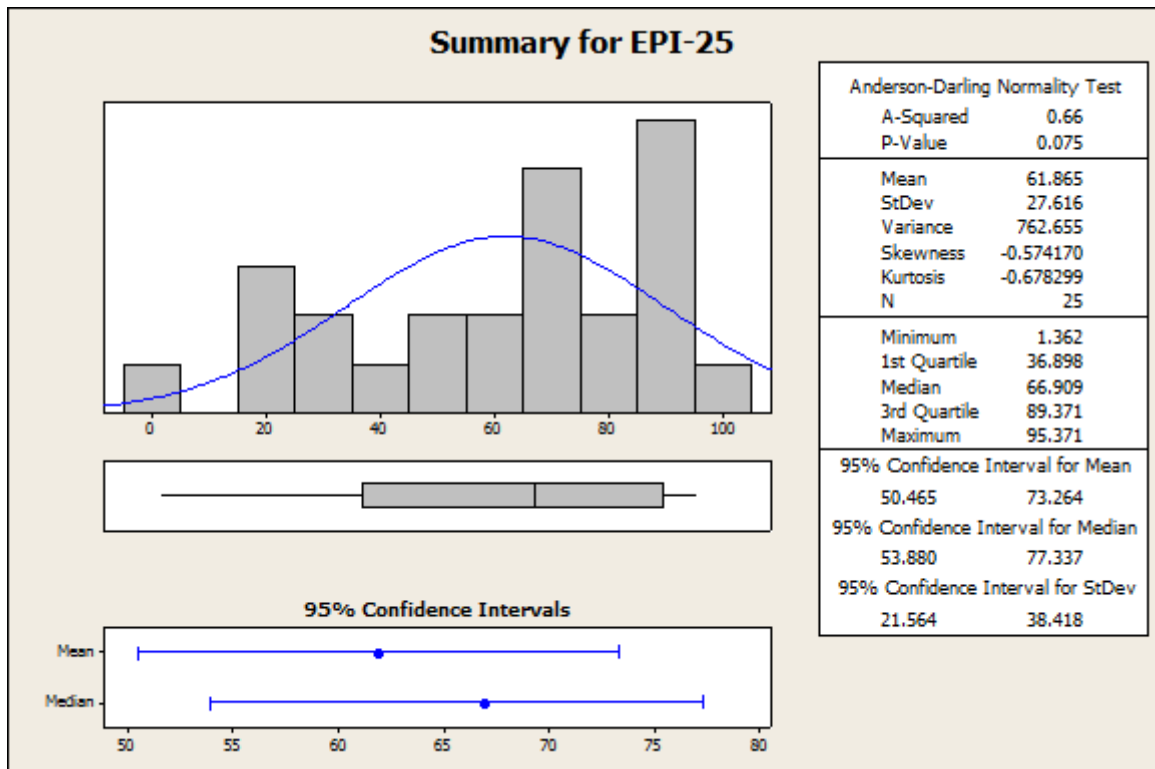
Obs: Visualmente, o teste de 2-Sample apresentou resultado muito próximo da ANOVA. Aqui a Letônia (Latvia) também figura como *outlier*.

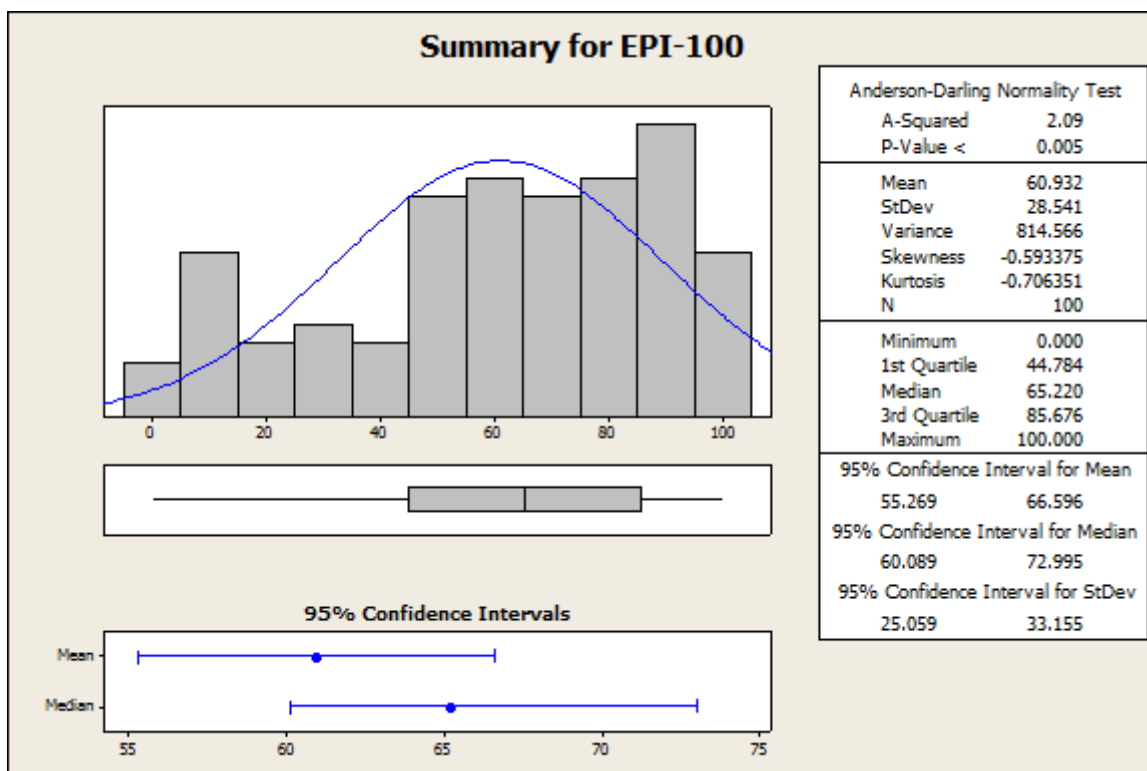
6.4 EPI-2014 – Índice de Performance Ambiental

Nas sessões que seguem, serão demonstrados os resultados da análise exploratória, análise de variância e teste de amostragem entre duas regiões para a variável EPI-2014.

6.4.1 Análise Exploratória

Abaixo os gráficos boxplots de análise exploratória com as amostragens de 25, 50 e 100 países, escolhidas aleatoriamente entre as 3 (três) regiões.





6.4.2 Análise de variância (ANOVA unstacked)

Os resultados para a ANOVA, considerando as amostras com 25, 50, 100 e 132 países, são demonstrados abaixo.

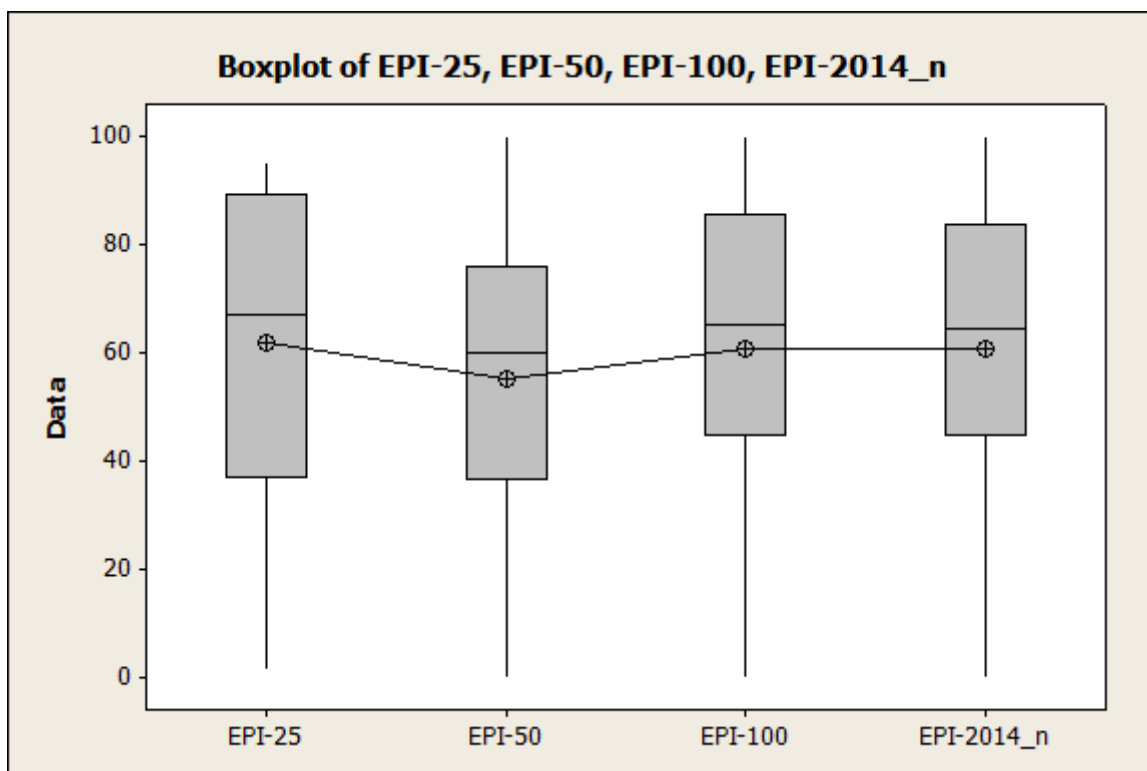
One-way ANOVA: EPI-25, EPI-50, EPI-100, EPI-2014_n

Source	DF	SS	MS	F	P
Factor	3	1454	485	0.63	0.598
Error	303	234390	774		
Total	306	235844			

S = 27.81 R-Sq = 0.62% R-Sq(adj) = 0.00%

Level	N	Mean	StDev	Individual 95% CIs For Mean Based on Pooled StDev
EPI-25	25	61.86	27.62	(-----*-----)
EPI-50	50	55.11	27.47	(-----*-----)
EPI-100	100	60.93	28.54	(-----*-----)
EPI-2014_n	132	60.81	27.42	(-----*-----)

Pooled StDev = 27.81



Obs: No gráfico acima, a amostra com 100 países apresentou o menor intervalo, denotando representar melhor a totalidade da população, com 95% de confiança.

Comparação entre as regiões AIBER e AVECO (ANOVA unstacked)

One-way ANOVA: epi-AIBER, epi-AVECO

Source	DF	SS	MS	F	P
Factor	1	4371.7	4371.7	44.48	0.000
Error	46	4520.9	98.3		
Total	47	8892.6			

S = 9.914 R-Sq = 49.16% R-Sq(adj) = 48.06%

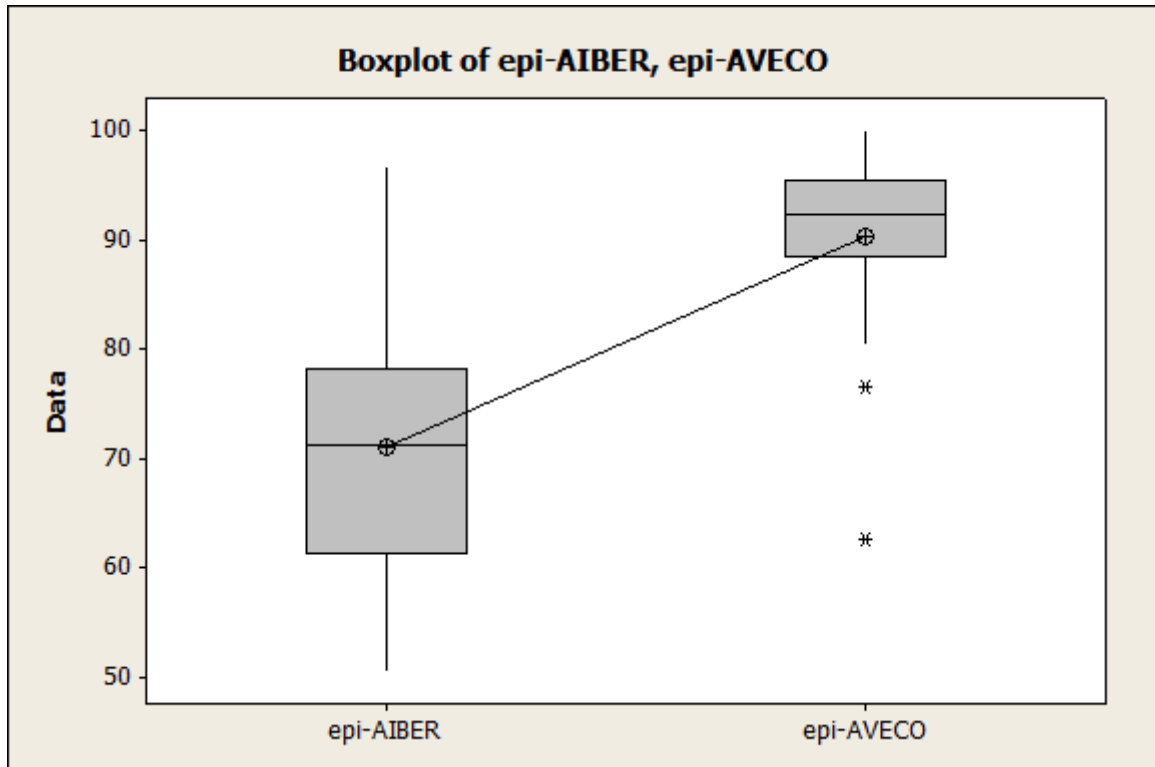
Level	N	Mean	StDev
epi-AIBER	21	71.054	11.913
epi-AVECO	27	90.292	8.045

Individual 95% CIs For Mean Based on Pooled StDev

Level	Lower CI	Upper CI
epi-AIBER	50.0	92.0
epi-AVECO	82.0	98.0

70.0 77.0 84.0 91.0

Pooled StDev = 9.914



Obs: Dois países do agrupamento AVECO se mostraram como *outliers*, são eles Israel e Republic of Korea.

6.4.3.3 Testes de amostragens (two-sample)

Teste de amostragens (T-Test).

Two-Sample T-Test and CI: epi-AIBER, epi-AVECO

Two-sample T for epi-AIBER vs epi-AVECO

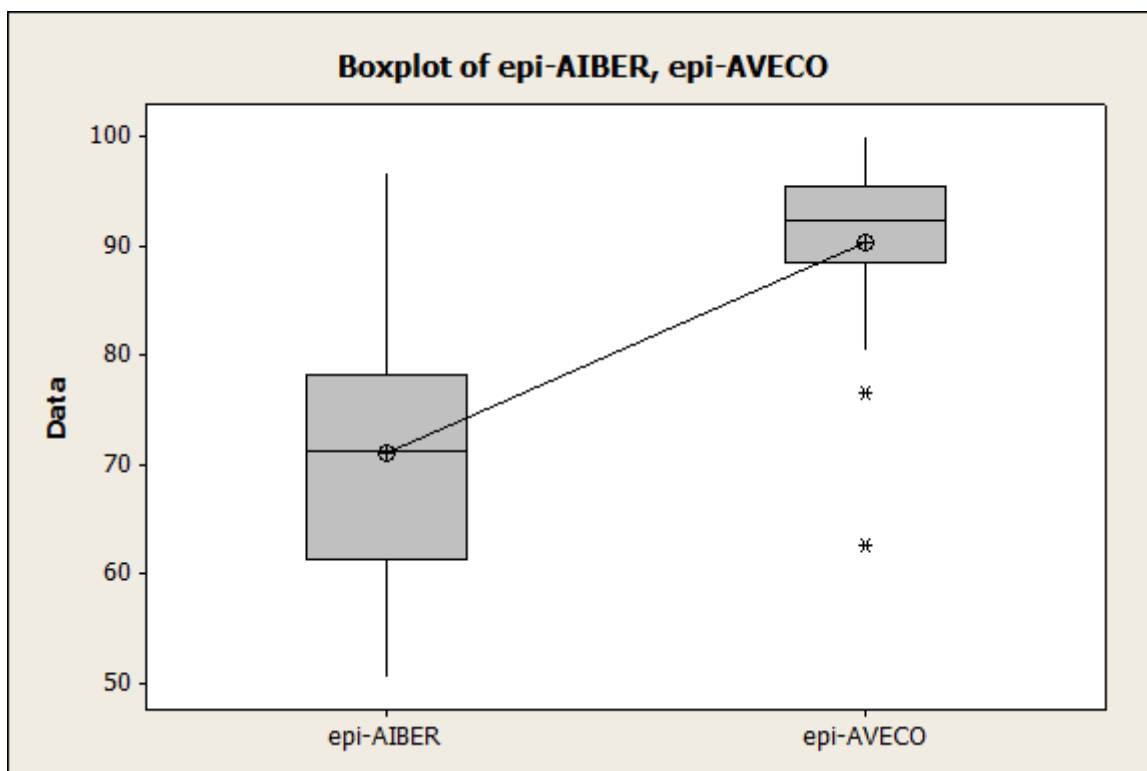
	N	Mean	StDev	SE Mean
epi-AIBER	21	71.1	11.9	2.6
epi-AVECO	27	90.29	8.04	1.5

Difference = μ (epi-AIBER) - μ (epi-AVECO)

Estimate for difference: -19.24

95% upper bound for difference: -14.12

T-Test of difference = 0 (vs <): T-Value = -6.36 P-Value = 0.000 DF = 33



Obs: O teste de “2-Sample t ” se apresentou praticamente idêntico ao boxplot da sessão anterior com ANOVA (unstacked). Inclusive repetindo Israel e Republic of Korea como *outliers* .

6.5. CONSIDERAÇÕES

O objetivo central deste trabalho foi comparar a região dos países Ibero-Americanos – AIBER, com a região AVECO – Economias Avançadas, em relação às variáveis SPI, IDH e EPI, todas do ano 2014. Para apoiar estas conclusões foram efetuados análises exploratórias dos dados, pesquisas por amostragens, análise de variância (ANOVA), testes de amostragens (two-sample) e gráficos boxplots. Como resultado, obtivemos que a região AVECO apresenta melhores índices de performance que a região AIBER.

Como conclusão secundária, para as pesquisas de amostragens realizadas em agrupamentos de 25, 50 e 100 indivíduos (países), nos casos estudados, a amostra com 100 países apresentou menor intervalo, portanto, representando melhor o comportamento do todo.

7. COMPONENTES PRINCIPAIS

Esta sessão se dedica a apresentar os resultados de análise das correlações e dos componentes principais (*principal components analysis*). Para tal, iniciamos com a análise das correlações e dendrogramas (*dendrogram*). Em seguida partimos para os componentes principais, análise e criação de novos índices .

Importante: Os índices SPI, IDH e EPI foram atualizados para o ano base de 2014. E os valores da variável IDH para os países Suíça, Suécia e Swaziland também foram corrigidos.

7.1. ANÁLISE DE COMPONENTES PRINCIPAIS, DENDROGRAMAS E CORRELAÇÕES

Nesta sessão faremos as análises de componentes principais, dendrogramas (matriz de dissimilaridade) e as correlações e regressões *stepwise*.

7.1.1 ANÁLISES DE COMPONENTES PRINCIPAIS

Nesta primeira análise foram consideradas todas as variáveis deste estudo (total de 30), abaixo os resultados da matrix de correlação e as variáveis e seus gráficos.

Results for: METODOS_new SPI HDI and EPI_version 2014_MINITAB.MTW

Principal Component Analysis: SPI-2014_n, Basic Human , Foundations , Opportuni

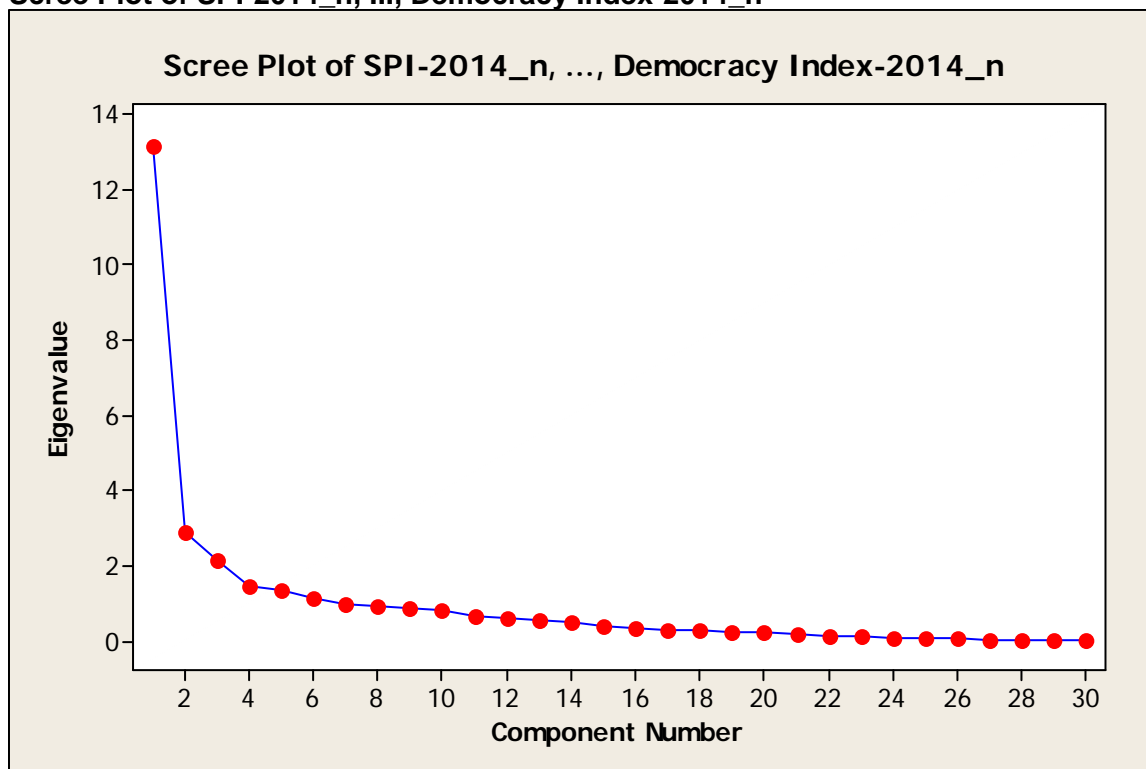
Eigenanalysis of the Correlation Matrix

Eigenvalue	13.148	2.856	2.117	1.419	1.344	1.110	0.964	0.897	0.841
Proportion	0.438	0.095	0.071	0.047	0.045	0.037	0.032	0.030	0.028
Cumulative	0.438	0.533	0.604	0.651	0.696	0.733	0.765	0.795	0.823
Eigenvalue	0.794	0.627	0.581	0.515	0.467	0.373	0.319	0.294	0.253
Proportion	0.026	0.021	0.019	0.017	0.016	0.012	0.011	0.010	0.008
Cumulative	0.850	0.871	0.890	0.907	0.923	0.935	0.946	0.955	0.964
Eigenvalue	0.224	0.203	0.181	0.139	0.100	0.079	0.059	0.038	0.025
Proportion	0.007	0.007	0.006	0.005	0.003	0.003	0.002	0.001	0.001
Cumulative	0.971	0.978	0.984	0.989	0.992	0.995	0.997	0.998	0.999
Eigenvalue	0.020	0.014	0.000						
Proportion	0.001	0.000	0.000						
Cumulative	1.000	1.000	1.000						

Variable	PC1	PC2	PC3	PC4
SPI-2014_n	0.272	-0.002	-0.009	-0.006
Basic Human Needs_n	0.250	0.163	-0.004	0.023
Foundations of Wellbeing_n	0.253	-0.079	0.045	0.006
Opportunity_n	0.259	-0.145	-0.054	-0.050
Health and Wellness_n	0.119	-0.245	0.199	0.505
Personal Freedom and Choice_n	0.245	-0.142	0.047	0.015
Greenhouse gas emissions_np	0.097	0.085	0.131	0.207
Water withdrawals_np	-0.048	-0.311	-0.166	-0.156
Biodiversity and habitat_n	0.093	-0.276	0.025	-0.005
Modern slavery_np	0.172	0.084	0.174	-0.037

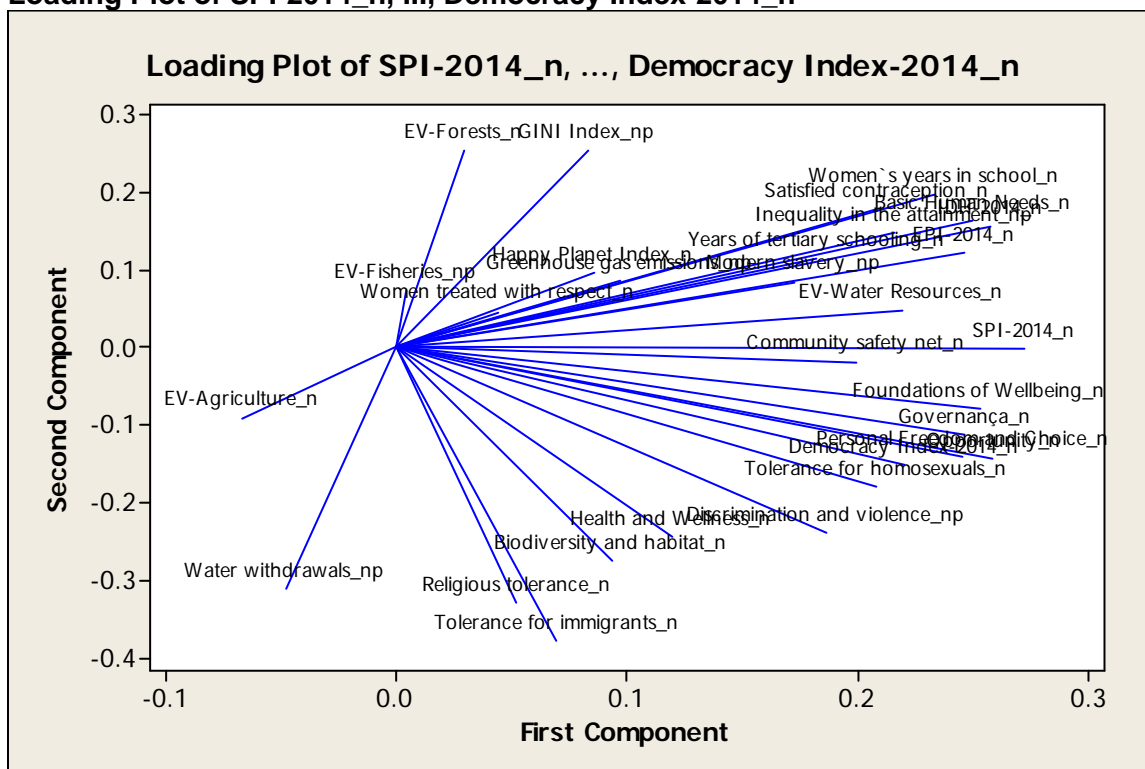
Satisfied contraception_n	0.209	0.179	0.226	-0.132
Women treated with respect_n	0.044	0.045	-0.405	0.465
Tolerance for immigrants_n	0.070	-0.380	-0.139	0.033
Tolerance for homosexuals_n	0.208	-0.181	0.067	0.122
Discrimination and violence_np	0.186	-0.239	-0.170	-0.149
Religious tolerance_n	0.052	-0.330	-0.009	-0.370
Community safety net_n	0.200	-0.019	0.077	-0.023
Years of tertiary schooling_n	0.182	0.113	-0.081	0.040
Women`s years in school_n	0.233	0.197	0.009	-0.221
Inequality in the attainment_np	0.216	0.148	-0.107	-0.239
IDH-2014_n	0.258	0.155	0.016	-0.007
Governança_n	0.246	-0.113	-0.161	0.071
EPI-2014_n	0.246	0.121	0.045	-0.139
EV-Water Resources_n	0.219	0.046	-0.190	0.175
EV-Agriculture_n	-0.067	-0.092	0.184	-0.080
EV-Forests_n	0.029	0.253	-0.194	-0.163
EV-Fisheries_np	0.004	0.071	-0.353	-0.094
Happy Planet Index_n	0.085	0.096	0.495	0.147
GINI Index_np	0.083	0.253	-0.283	0.196
Democracy Index-2014_n	0.220	-0.152	0.015	0.003

Scree Plot of SPI-2014_n, ..., Democracy Index-2014_n



O gráfico mostra visualmente o que vimos mais acima na matriz de correlação, onde denota que a variável SPI explica 43.8% dos resultados e se por convenção considerarmos as componentes com contribuição igual ou maior que 1, teríamos as 6 (seis) primeiras variáveis explicando 73.3% da variabilidade dos dados.

Loading Plot of SPI-2014_n, ..., Democracy Index-2014_n



Este gráfico mostra uma expressiva concentração dos dados na parte superior direita, onde as componentes apresentaram resultados positivos. Apenas duas variáveis apresentaram valores negativos nos dois eixos.

Conforme, Las Casas e Guevara (2010)², por convenção somente deve ser consideradas as componentes com uma contribuição igual ou superior a 1. Partindo dessa convenção, listamos abaixo as 6 (seis) variáveis as quais considerando o acumulado, explicam 73.3% da variabilidade dos dados.

Variable	PC1	PC2	PC3	PC4
SPI-2014_n	0.272	-0.002	-0.009	-0.006
Basic Human Needs_n	0.250	0.163	-0.004	0.023
Foundations of Wellbeing_n	0.253	-0.079	0.045	0.006
Opportunity_n	0.259	-0.145	-0.054	-0.050
Health and Wellness_n	0.119	-0.245	0.199	0.505
Personal Freedom and Choice_n	0.245	-0.142	0.047	0.015

² LAS CASAS, Alexandre Luzzi; GUEVARA, Arnoldo José de Hoyos. *Pesquisa de Marketing*. São Paulo: Atlas, 2010.

7.2 DENDROGRAMA (Dendrogram)

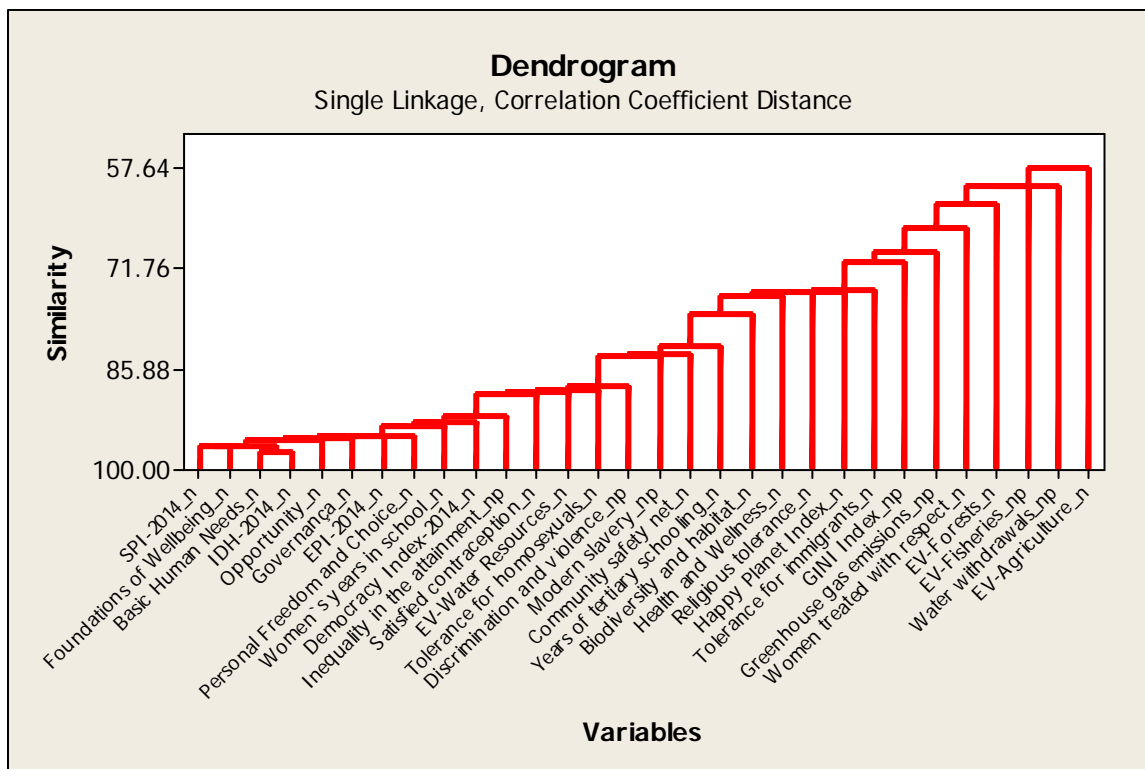
Abaixo os resultados das análises de *clusters*, considerando o grau de similaridade entre as 30 variáveis deste estudo.

Cluster Analysis of Variables: SPI-2014, Basic Human Needs, Foundations of Wellbeing, ...

Correlation Coefficient Distance, Single Linkage
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	29	97.8566	0.042868	2	21	2
2	28	96.9544	0.060912	1	3	2
3	27	96.8730	0.062539	1	2	4
4	26	96.0972	0.078056	1	4	5
5	25	95.8063	0.083875	1	22	6
6	24	95.4891	0.090218	1	23	7
7	23	95.3580	0.092840	1	6	8
8	22	94.2163	0.115674	1	19	9
9	21	93.4716	0.130567	1	30	10
10	20	92.5817	0.148366	1	20	11
11	19	89.6856	0.206288	1	11	12
12	18	89.1677	0.216646	1	24	13
13	17	89.0470	0.219061	1	14	14
14	16	88.3252	0.233497	1	15	15
15	15	84.0927	0.318146	1	10	16
16	14	84.0626	0.318747	1	17	17
17	13	82.9000	0.342001	1	18	18
18	12	78.2833	0.434334	1	9	19
19	11	75.8883	0.482233	1	5	20
20	10	75.2897	0.494205	1	16	21
21	9	75.0977	0.498045	1	28	22
22	8	74.9784	0.500432	1	13	23
23	7	71.0320	0.579361	1	29	24
24	6	69.6158	0.607684	1	7	25
25	5	66.3052	0.673897	1	12	26
26	4	62.8294	0.743413	1	26	27
27	3	60.2922	0.794155	1	27	28
28	2	60.1415	0.797170	1	8	29
29	1	57.6359	0.847283	1	25	30

Dendrogram



O gráfico de Dendrograma (matriz de dissimilaridade) reforça os resultados da tabela acima de análise de clusters, onde seriam necessários 29 clusters para ter similaridade de 97.85%, 13 para chegar a 82.90% e 7 clusters para atingir 71.03% de similaridade entre as variáveis.

7.3 CORRELAÇÕES

As correlações abaixo apresentaram resultados não significativos ($p > 0.05$). E os coeficientes de correlação de Pearson com valores próximos a 0 (zero) indica uma relação fraca ou inexistente entre as duas variáveis.

Correlations: PC1, PC2, PC3, PC4

	PC1	PC2	PC3
PC2	0.000 1.000		
PC3	0.000 1.000	-0.000 1.000	
PC4	0.000 1.000	-0.000 1.000	0.000 1.000

Cell Contents: Pearson correlation
P-Value

7.4 Regressões Stepwise

Nesta sessão apresentamos os resultados da análise de regressão *stepwise*, utilizando como referência cada variável PC (principal componente) criada, contra as outras 30 variáveis deste trabalho como preditoras.

Stepwise Regression: PC1 versus SPI-2014_n, Basic Human Needs_n, ...

Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15

Response is PC1 on 30 predictors, with N = 132

Step	1	2	3	4	5	6
Constant	-7.832	-7.737	-7.874	-9.449	-10.026	-9.872
SPI-2014_n	0.1402	0.1162	0.0572	0.0598	0.0620	0.0631
T-Value	70.42	25.14	6.89	7.95	9.09	10.16
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
Opportunity_n		0.0272	0.0500	0.0470	0.0416	0.0375
T-Value		5.64	10.25	10.58	10.04	9.74
P-Value		0.000	0.000	0.000	0.000	0.000
IDH-2014_n			0.0401	0.0347	0.0309	0.0260
T-Value			7.99	7.47	7.26	6.53
P-Value			0.000	0.000	0.000	0.000
Modern slavery_np				0.0209	0.0211	0.0215
T-Value				5.42	6.04	6.76
P-Value				0.000	0.000	0.000
Community safety net_n					0.0126	0.0130
T-Value					5.40	6.15
P-Value					0.000	0.000
EV-Water Resources_n						0.0071
T-Value						5.22
P-Value						0.000
S	0.582	0.523	0.429	0.388	0.351	0.319
R-Sq	97.45	97.95	98.63	98.89	99.10	99.26
R-Sq(adj)	97.43	97.92	98.60	98.85	99.06	99.22

Para o PC1, temos as variáveis SPI-2014, Opportunity e IDH-2014 que explicam 98.63% dos dados.

Stepwise Regression: PC2 versus SPI-2014_n, Basic Human Needs_n, ...

Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15

Response is PC2 on 30 predictors, with N = 132

Step	1	2	3	4	5
Constant	2.512	1.529	2.514	1.056	1.705

Tolerance for immigrants_n	-0.0485	-0.0501	-0.0448	-0.0419	-0.0363
T-Value	-9.54	-12.24	-12.46	-13.63	-13.29
P-Value	0.000	0.000	0.000	0.000	0.000
EV-Forests_n		0.0269	0.0231	0.0200	0.0188
T-Value		8.48	8.33	8.36	9.19
P-Value		0.000	0.000	0.000	0.000
Water withdrawals_np			-0.0181	-0.0185	-0.0170
T-Value			-6.89	-8.30	-8.94
P-Value			0.000	0.000	0.000
GINI Index_np				0.0231	0.0200
T-Value				7.18	7.20
P-Value				0.000	0.000
Religious tolerance_n					-0.0132
T-Value					-7.03
P-Value					0.000
Health and Wellness_n					
T-Value					
P-Value					
S	1.30	1.05	0.897	0.760	0.647
R-Sq	41.17	62.22	72.44	80.40	85.91
R-Sq(adj)	40.71	61.63	71.79	79.78	85.36

Para o PC2, as variáveis Tolerância a Imigrantes, Índice de Cobertura de Florestas e Extração de Água, explicam 72.44% dos dados.

Stepwise Regression: PC3 versus SPI-2014_n, Basic Human Needs_n, ...

Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15

Response is PC3 on 30 predictors, with N = 132

Step	1	2	3	4	5
Constant	-2.3539	-0.4865	1.3870	2.0243	2.1238
Happy Planet Index_n	0.0486	0.0414	0.0381	0.0387	0.0388
T-Value	11.82	12.35	13.71	16.14	17.44
P-Value	0.000	0.000	0.000	0.000	0.000
Women treated with respect_n		-0.0279	-0.0244	-0.0196	-0.0190
T-Value		-8.87	-9.33	-8.24	-8.61
P-Value		0.000	0.000	0.000	0.000
EV-Fisheries_np			-0.0268	-0.0252	-0.0242
T-Value			-8.02	-8.70	-8.97
P-Value			0.000	0.000	0.000
GINI Index_np				-0.0166	-0.0151
T-Value				-6.68	-6.50
P-Value				0.000	0.000
EV-Forests_n					-0.0076
T-Value					-4.69
P-Value					0.000
Water withdrawals_np					
T-Value					
P-Value					

S	1.01	0.802	0.657	0.568	0.526
R-Sq	51.82	70.06	80.07	85.25	87.44
R-Sq(adj)	51.45	69.60	79.60	84.78	86.94

Para o PC3, temos as variáveis Índice de Felicidade do Planeta, Índice de Respeito à Mulher e Extração de Peixes na Costa, as quais explicam 80.07% dos dados.

Stepwise Regression: PC4 versus SPI-2014_n, Basic Human Needs_n, ...

Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15

Response is PC4 on 30 predictors, with N = 132

Step	1	2	3	4	5
Constant	-2.319	-1.436	-2.381	-1.787	-4.187
Health and Wellness_n	0.0343	0.0377	0.0334	0.0335	0.0311
T-Value	8.58	12.20	13.51	18.05	19.95
P-Value	0.000	0.000	0.000	0.000	0.000
Religious tolerance_n		-0.01880	-0.01664	-0.01497	-0.01500
T-Value		-9.54	-10.63	-12.63	-15.34
P-Value		0.000	0.000	0.000	0.000
Women treated with respect_n			0.0204	0.0231	0.0232
T-Value			9.02	13.45	16.45
P-Value			0.000	0.000	0.000
Inequality in the attainment_np				-0.0135	-0.0152
T-Value				-10.04	-13.45
P-Value				0.000	0.000
Greenhouse gas emissions_np					0.0275
T-Value					7.79
P-Value					0.000
GINI Index_np					
T-Value					
P-Value					
S	0.955	0.735	0.576	0.432	0.356
R-Sq	36.16	62.56	77.12	87.24	91.39
R-Sq(adj)	35.66	61.98	76.58	86.84	91.05

Para o PC4, as variáveis Saúde e Bem-Estar, Tolerância à Religião e Índice de Respeito à Mulher, explicam 77.12% dos dados.

7.5 Criando indicadores a partir dos PC1, PC2, PC3 e PC4

Aproveitamos os resultados da sessão anterior, que através da análise de regressão *stepwise*, relacionamos cada componente principal (PC), com as 30 variáveis afim de obter a mais alta correlação com as variáveis de resposta.

Abaixo, listamos as 3 (três) variáveis que apresentaram esta correlação, uma sugestão para o nome do indicador proposto e uma sugestão para o nome da variável.

PC1 = SPI + Opportunity + IDH

Nome do Indicador = Growth or Development (crescimento ou desenvolvimento) = socoppi

PC2 = Immigrants Tol. + EV-Forests + Water withdrawals

Nome do Indicador = Careful (consciência ou atento) = immigepi

PC3 = HPI + Women treated w/respect + EV-Fisheries

Nome do Indicador = Happiness (Felicidade) = hpiwfish

PC4 = Health & Wellness + Religious Tol. + Women treated w/respect

Nome do Indicador = Well-being or Take care (cuidado ou estar bem) = wellreliwo

7.6. CONSIDERAÇÕES

Nesta sessão foram realizadas análises de componentes principais para a construção de índices com alto poder explicativo, utilizando o menor número possível de variáveis.

De acordo com as análises realizadas, notamos que o agrupamento de variáveis é pertinente, porém, por alguma característica dos dados, aparentemente há um viés mais social do que natural (ecológico) nos indicadores gerados, ou seja, das 11 variáveis que se despontaram na análise, temos 8 delas ligadas a índices que medem dimensões ligadas ao indivíduo ou seu ambiente social e apenas 3 variáveis representando o uso de algum recurso natural do meio ambiente.

A variável *Biodiversity and Habitat* por exemplo, não se destacou em nenhuma das análises aqui realizadas.

8. ANÁLISE DE CONGLOMERADOS

O objetivo desta sessão é efetuar análise de conglomerados, agrupar dados de acordo com sua similaridade a partir de pontos definidos pelas médias das variáveis para os indivíduos de cada agrupamento (centróides). Será realizado também análise de dendrogramas e clusters. Portanto, iniciamos com a análise dos agrupamentos e dendrogramas

(*dendrogram*). Em seguida partimos para categorização e separação por região, análises com a nova amostra e gráficos de dispersão (*scatterplot 3D*).

8.1. METODOLOGIA

Nesta sessão faremos alguns ensaios (análise multivariada) a fim de identificar agrupamentos por grau de similaridade e a partir de então a amostra será categorizada e separada por região, para que possamos comparar AIBER e AVECO. A saber, serão considerados para este estudo somente países integrantes das regiões AIBER e AVECO, os demais países (região OTHERS) não serão contemplados nas análises.

As variáveis para este estudo, serão: EPI-2014, Governança, Biodiversity and habitat, Democracy Index-2014, SPI-2014, Opportunity, IDH-2014, Tolerance for immigrants, EV-Forests, Water withdrawals, HPI, Women treated with respect, EV-Fisheries, Health and Wellness, Religious tolerance. E os índices PC1, PC2, PC3 e PC4, gerados a partir da funcionalidade de components principais. Estes PCs foram gerados considerando as 30 variáveis referentes ao tema, como preditoras e apresentaram os seguintes resultados em sua composição:

Para o PC1, as variáveis SPI-2014, Opportunity e IDH-2014 que explicam 98.63% dos dados.

Para o PC2, Tolerância a Imigrantes, Índice de Cobertura de Florestas e Extração de Água, explicam 72.44% dos dados.

Para o PC3, as variáveis Índice de Felicidade do Planeta, Índice de Respeito à Mulher e Extração de Peixes na Costa, as quais explicam 80.07% dos dados.

E para o PC4, Saúde e Bem-Estar, Tolerância à Religião e Índice de Respeito à Mulher, explicam 77.12% dos dados.

Em tempo, optou-se em adicionar a este trabalho as variáveis Governança, Democracy Index-2014 e Biodiversity and Habitat, sendo as duas primeiras devido ao forte viés social nos índices PCs, e a terceira por fazer parte desta temática mas que não apresentou relevância até o momento nas análises.

8.2 Dendrogramas e Análise de Clusters (por variáveis e por observações)

Para encontrar a melhor combinação de agrupamento para as variáveis deste estudo, serão realizadas análises com 1, 10, 20 e 30 clusters, abaixo estão listadas apenas algumas para exemplificação.

8.2.1 Análise com 20 Clusters (por variáveis e por observações)

Análise com 20 Clusters (por variáveis e por observações) e Dendrogramas

Cluster Analysis of Observations: EPI-2014_n, Governança_n, Biodiversity, ...

Euclidean Distance, Single Linkage
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	131	89.5016	28.1919	32 108	32	2
2	130	88.8134	30.0399	43 88	43	2
3	129	88.7094	30.3193	32 99	32	3
4	128	88.3161	31.3755	49 90	49	2
5	127	87.2973	34.1112	11 127	11	2
6	126	86.7834	35.4914	46 49	46	3
7	125	86.0066	37.5771	7 93	7	2
8	124	85.7307	38.3182	18 91	18	2
9	123	85.6685	38.4851	7 115	7	3
10	122	85.3638	39.3034	43 126	43	3
11	121	85.0209	40.2243	21 119	21	2
12	120	84.7665	40.9073	116 129	116	2
13	119	84.4339	41.8006	21 28	21	3
14	118	84.3370	42.0607	41 43	41	4
15	117	84.0043	42.9541	7 40	7	4
16	116	83.9718	43.0414	13 46	13	4
17	115	83.8607	43.3397	7 114	7	5
18	114	83.8089	43.4789	12 18	12	3
19	113	83.5603	44.1464	19 21	19	4
20	112	83.4068	44.5586	63 92	63	2
21	111	83.3470	44.7193	7 33	7	6
22	110	83.3451	44.7244	68 72	68	2
23	109	83.2021	45.1083	13 36	13	5
24	108	83.0136	45.6146	13 97	13	6
25	107	83.0057	45.6358	13 35	13	7
26	106	82.6211	46.6685	5 8	5	2
27	105	82.2042	47.7881	7 32	7	9
28	104	82.1788	47.8562	60 110	60	2
29	103	81.9167	48.5602	13 38	13	8
30	102	81.8817	48.6541	22 51	22	2
31	101	81.8590	48.7150	75 132	75	2
32	100	81.8290	48.7956	7 41	7	13
33	99	81.8213	48.8163	7 30	7	14
34	98	81.8149	48.8335	7 22	7	16
35	97	81.7963	48.8833	19 78	19	5
36	96	81.6484	49.2806	53 118	53	2
37	95	81.6243	49.3452	117 123	117	2
38	94	81.5449	49.5585	4 16	4	2
39	93	81.3389	50.1117	7 107	7	17
40	92	81.1378	50.6517	58 60	58	3
41	91	81.0669	50.8420	12 105	12	4
42	90	81.0510	50.8848	17 83	17	2
43	89	81.0269	50.9495	45 80	45	2
44	88	80.9641	51.1181	47 63	47	3
45	87	80.8449	51.4381	75 117	75	4
46	86	80.7638	51.6559	12 75	12	8
47	85	80.7365	51.7294	15 86	15	2
48	84	80.7012	51.8242	14 81	14	2
49	83	80.5688	52.1798	4 27	4	3
50	82	80.3391	52.7965	13 95	13	9
51	81	80.3181	52.8530	61 66	61	2
52	80	80.2338	53.0792	94 131	94	2
53	79	80.1247	53.3723	7 11	7	19
54	78	80.0838	53.4820	12 47	12	11
55	77	80.0127	53.6729	19 103	19	6
56	76	79.6069	54.7628	102 124	102	2

57	75	79.5361	54.9527	71	77	71	2
58	74	79.4987	55.0532	4	13	4	12
59	73	78.5698	57.5477	7	101	7	20
60	72	78.5578	57.5798	3	12	3	12
61	71	78.5390	57.6303	3	112	3	13
62	70	78.5047	57.7224	4	29	4	13
63	69	78.4412	57.8930	65	104	65	2
64	68	78.3884	58.0349	3	85	3	14
65	67	78.3749	58.0710	3	19	3	20
66	66	78.3673	58.0916	52	94	52	3
67	65	78.2565	58.3891	61	69	61	3
68	64	78.2292	58.4623	64	100	64	2
69	63	78.1194	58.7571	3	24	3	21
70	62	78.1078	58.7883	7	58	7	23
71	61	78.0899	58.8365	7	64	7	25
72	60	78.0445	58.9583	2	122	2	2
73	59	78.0431	58.9622	5	26	5	3
74	58	77.9357	59.2505	7	68	7	27
75	57	77.9085	59.3235	45	98	45	3
76	56	77.7467	59.7579	7	45	7	30
77	55	77.4969	60.4289	2	61	2	5
78	54	77.4808	60.4721	7	17	7	32
79	53	77.4142	60.6510	84	121	84	2
80	52	77.2200	61.1723	2	7	2	37
81	51	77.1776	61.2864	5	54	5	4
82	50	77.0991	61.4970	2	53	2	39
83	49	76.8489	62.1689	5	84	5	6
84	48	76.8296	62.2207	2	50	2	40
85	47	76.7724	62.3743	2	5	2	46
86	46	76.7586	62.4114	2	130	2	47
87	45	76.7416	62.4571	2	31	2	48
88	44	76.6390	62.7326	65	125	65	3
89	43	76.6126	62.8034	4	96	4	14
90	42	76.5211	63.0491	2	39	2	49
91	41	76.4917	63.1280	2	37	2	50
92	40	76.4707	63.1846	2	4	2	64
93	39	76.3227	63.5820	1	2	1	65
94	38	76.1063	64.1630	1	76	1	66
95	37	76.0938	64.1966	1	15	1	68
96	36	75.5654	65.6156	6	89	6	2
97	35	75.5221	65.7319	1	128	1	69
98	34	75.3933	66.0778	82	120	82	2
99	33	75.2795	66.3832	1	42	1	70
100	32	75.2023	66.5906	1	116	1	72

Final Partition

Number of clusters: 20

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	85	695709	88.6533	154.971
Cluster2	25	100090	62.5368	85.858
Cluster3	3	5129	41.1018	44.547
Cluster4	1	0	0.0000	0.000
Cluster5	1	0	0.0000	0.000
Cluster6	1	0	0.0000	0.000
Cluster7	2	2354	34.3084	34.308
Cluster8	1	0	0.0000	0.000
Cluster9	1	0	0.0000	0.000
Cluster10	1	0	0.0000	0.000
Cluster11	1	0	0.0000	0.000
Cluster12	1	0	0.0000	0.000
Cluster13	1	0	0.0000	0.000
Cluster14	1	0	0.0000	0.000
Cluster15	1	0	0.0000	0.000
Cluster16	1	0	0.0000	0.000

Cluster17	1	0	0.0000	0.000
Cluster18	1	0	0.0000	0.000
Cluster19	2	1499	27.3814	27.381
Cluster20	1	0	0.0000	0.000

Cluster Centroids

Variable	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
EPI-2014_n	73.1199	23.7148	93.642	8.716	84.360
Governana_n	52.3117	24.0948	92.669	19.989	21.841
Biodiversity and habitat_n	64.2159	57.5424	59.290	39.680	41.250
Democracy Index-2014_n	58.5475	32.5640	88.942	50.829	26.066
Opportunity_n	53.3460	22.1821	95.892	23.752	31.115
IDH-2014_n	71.7066	19.5627	96.215	37.246	75.570
Tolerance for immigrants_n	49.6696	61.7534	90.411	50.685	35.616
EV-Forests_n	40.8304	28.1260	100.000	22.830	68.230
Water withdrawals_np	54.6389	93.0493	48.131	87.049	88.172
HPI_n	54.5616	24.7487	54.824	81.315	35.767
Women treated with respect_n	53.0723	55.3171	82.114	78.049	60.976
EV-Fisheries_np	70.4019	70.0554	77.959	100.000	71.024
Health and Wellness_n	72.1929	59.9820	90.859	79.958	19.892
Religious tolerance_n	60.0000	57.3333	77.778	0.000	100.000
SPI-2014_n	66.9087	21.6542	96.208	34.939	58.591

Variable	Cluster6	Cluster7	Cluster8	Cluster9	Cluster10
EPI-2014_n	26.390	29.6006	63.531	50.1493	76.596
Governana_n	24.541	27.7585	34.564	7.3225	65.156
Biodiversity and habitat_n	78.930	16.5700	18.660	0.0000	39.340
Democracy Index-2014_n	38.981	18.3649	52.370	32.4645	72.749
Opportunity_n	26.719	24.6075	45.589	4.7724	51.115
IDH-2014_n	34.670	25.5671	48.262	51.3087	91.672
Tolerance for immigrants_n	6.849	48.6301	49.315	23.2877	16.438
EV-Forests_n	0.000	46.0350	63.120	39.7200	39.720
Water withdrawals_np	91.154	35.0405	64.413	30.3363	3.492
HPI_n	42.784	41.8981	68.954	64.1798	78.689
Women treated with respect_n	91.463	64.6341	54.354	15.8537	53.659
EV-Fisheries_np	100.000	85.5118	74.708	71.0235	100.000
Health and Wellness_n	83.858	36.4686	39.064	54.7555	83.228
Religious tolerance_n	100.000	83.3333	100.000	0.0000	0.000
SPI-2014_n	34.669	26.6175	49.353	21.9986	69.734

Variable	Cluster11	Cluster12	Cluster13	Cluster14
EPI-2014_n	74.505	67.5439	24.6174	18.794
Governana_n	46.453	26.4472	22.8075	44.183
Biodiversity and habitat_n	88.040	14.1100	93.8500	1.300
Democracy Index-2014_n	69.905	19.9052	8.5308	61.256
Opportunity_n	64.819	37.9278	19.3721	44.364
IDH-2014_n	62.166	73.8266	38.0762	24.895
Tolerance for immigrants_n	73.973	43.8356	58.9041	32.877
EV-Forests_n	18.520	74.8200	13.2800	39.720
Water withdrawals_np	0.000	19.6290	99.7169	20.547
HPI_n	86.724	29.2259	64.0348	48.454
Women treated with respect_n	14.634	68.2927	86.5854	53.659
EV-Fisheries_np	100.000	71.0235	71.0235	71.024
Health and Wellness_n	81.908	0.0000	74.8275	44.614
Religious tolerance_n	100.000	66.6667	66.6667	100.000
SPI-2014_n	67.919	48.2926	35.6039	29.367

Variable	Cluster15	Cluster16	Cluster17	Cluster18
EPI-2014_n	76.3718	0.000	62.990	24.4681
Governana_n	44.7385	22.606	69.407	22.5662
Biodiversity and habitat_n	27.5400	37.080	43.700	61.7900
Democracy Index-2014_n	56.3981	34.716	79.147	38.8626
Opportunity_n	48.5400	26.515	61.068	20.6907
IDH-2014_n	66.9884	27.133	71.908	33.4578

Tolerance for immigrants_n	42.4658	46.575	80.822	54.7945
EV-Forests_n	25.3400	7.750	100.000	63.1200
Water withdrawals_np	47.0850	75.049	61.340	51.9583
HPI_n	13.7121	58.475	33.748	55.5695
Women treated with respect_n	57.3171	25.610	68.293	40.2439
EV-Fisheries_np	71.0235	100.000	23.915	71.0235
Health and Wellness_n	61.5662	78.938	80.408	71.5872
Religious tolerance_n	66.6667	66.667	66.667	33.3333
SPI-2014_n	64.2164	20.992	73.832	34.1121

Variable	Cluster19	Cluster20	Grand centroid
EPI-2014_n	83.0627	53.0982	60.8064
Governança_n	25.5564	36.2520	45.0180
Biodiversity and habitat_n	47.4250	67.7000	59.5132
Democracy Index-2014_n	34.5379	49.7630	51.9334
Opportunity_n	41.8289	29.0267	45.7947
IDH-2014_n	71.2168	68.5676	59.6370
Tolerance for immigrants_n	27.3973	32.8767	51.8057
EV-Forests_n	33.7950	33.7600	39.7221
Water withdrawals_np	66.6924	39.8425	61.3499
HPI_n	32.4759	64.6437	48.4542
Women treated with respect_n	34.1463	67.0732	54.3488
EV-Fisheries_np	69.5419	0.0000	71.0251
Health and Wellness_n	5.9256	74.0774	67.5908
Religious tolerance_n	16.6667	0.0000	59.3434
SPI-2014_n	54.3674	48.7239	55.8449

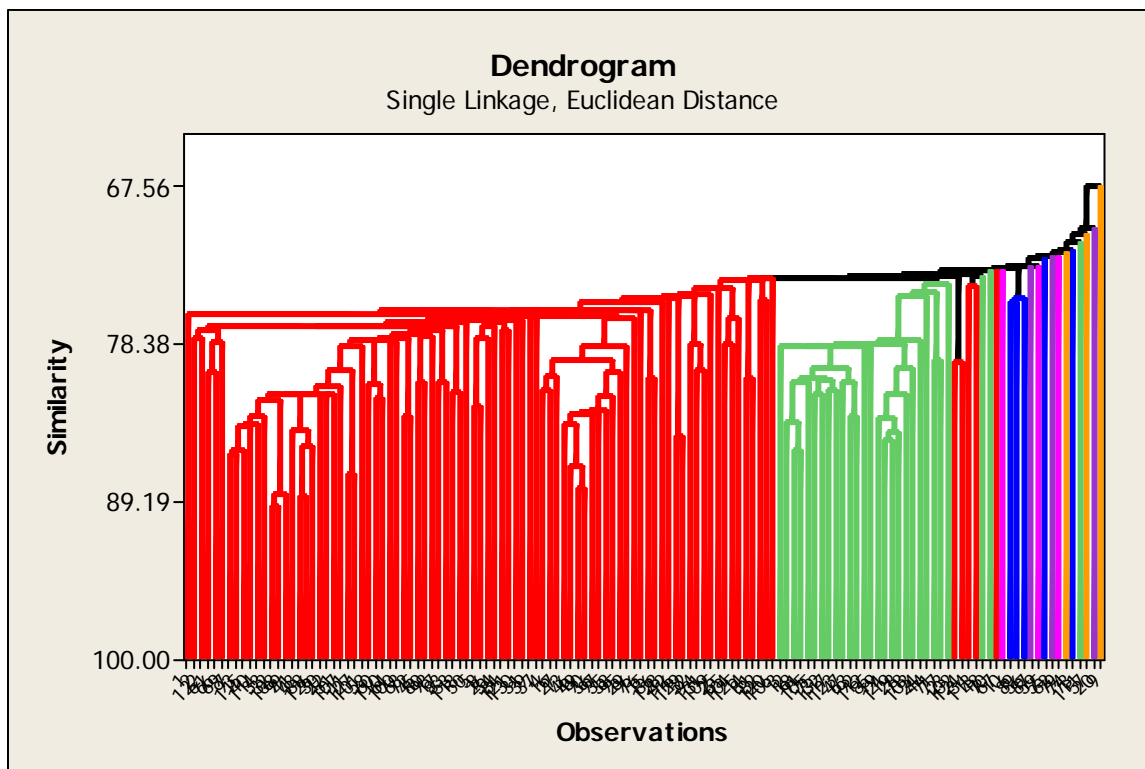
Distances Between Cluster Centroids

	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6	Cluster7
Cluster1	0.000	111.928	114.173	127.244	100.199	125.431	117.111
Cluster2	111.928	0.000	201.747	99.358	120.286	98.246	87.759
Cluster3	114.173	201.747	0.000	207.816	165.596	210.950	195.919
Cluster4	127.244	99.358	207.816	0.000	165.912	127.157	127.741
Cluster5	100.199	120.286	165.596	165.912	0.000	136.457	107.542
Cluster6	125.431	98.246	210.950	127.157	136.457	0.000	122.288
Cluster7	117.111	87.759	195.919	127.741	107.542	122.288	0.000
Cluster8	83.402	110.326	147.540	140.111	71.715	131.373	78.896
Cluster9	137.179	132.577	228.126	120.014	155.286	177.371	116.016
Cluster10	101.543	180.615	153.314	145.266	173.208	184.458	164.015
Cluster11	100.562	169.621	149.909	182.536	165.418	167.758	156.466
Cluster12	117.635	140.148	175.839	169.148	87.443	172.895	86.641
Cluster13	118.360	74.472	204.746	104.908	134.147	81.789	121.142
Cluster14	120.452	117.649	188.487	147.192	132.095	135.951	63.822
Cluster15	60.528	110.143	141.390	141.484	95.012	128.669	101.596
Cluster16	124.917	73.465	217.871	93.589	148.809	101.635	93.745
Cluster17	93.819	149.498	90.438	174.437	132.320	184.928	153.976
Cluster18	93.048	72.535	180.152	88.857	127.666	127.986	89.405
Cluster19	98.843	121.433	188.501	140.776	99.318	160.240	124.420
Cluster20	105.303	135.688	185.527	131.596	158.050	167.521	152.309

	Cluster8	Cluster9	Cluster10	Cluster11	Cluster12	Cluster13
Cluster1	83.402	137.179	101.543	100.562	117.635	118.360
Cluster2	110.326	132.577	180.615	169.621	140.148	74.472
Cluster3	147.540	228.126	153.314	149.909	175.839	204.746
Cluster4	140.111	120.014	145.266	182.536	169.148	104.908
Cluster5	71.715	155.286	173.208	165.418	87.443	134.147
Cluster6	131.373	177.371	184.458	167.758	172.895	81.789
Cluster7	78.896	116.016	164.015	156.466	86.641	121.142
Cluster8	0.000	135.056	149.349	131.931	92.010	132.277
Cluster9	135.056	0.000	131.635	182.030	128.476	165.763
Cluster10	149.349	131.635	0.000	138.408	151.909	187.218
Cluster11	131.931	182.030	138.408	0.000	170.197	168.610
Cluster12	92.010	128.476	151.909	170.197	0.000	166.436
Cluster13	132.277	165.763	187.218	168.610	166.436	0.000
Cluster14	81.415	132.133	159.060	147.580	111.344	153.847
Cluster15	86.080	130.287	119.344	130.094	98.453	138.216
Cluster16	117.932	121.437	170.247	152.934	160.328	101.466

Cluster17	114.685	182.459	156.109	160.908	141.031	168.718
Cluster18	104.226	97.677	137.606	143.928	125.100	100.478
Cluster19	115.447	108.923	133.128	167.404	97.502	147.341
Cluster20	148.101	124.851	127.559	172.857	150.406	138.286
	Cluster14	Cluster15	Cluster16	Cluster17	Cluster18	Cluster19
Cluster1	120.452	60.528	124.917	93.819	93.048	98.843
Cluster2	117.649	110.143	73.465	149.498	72.535	121.433
Cluster3	188.487	141.390	217.871	90.438	180.152	188.501
Cluster4	147.192	141.484	93.589	174.437	88.857	140.776
Cluster5	132.095	95.012	148.809	132.320	127.666	99.318
Cluster6	135.951	128.669	101.635	184.928	127.986	160.240
Cluster7	63.822	101.596	93.745	153.976	89.405	124.420
Cluster8	81.415	86.080	117.932	114.685	104.226	115.447
Cluster9	132.133	130.287	121.437	182.459	97.677	108.923
Cluster10	159.060	119.344	170.247	156.109	137.606	133.128
Cluster11	147.580	130.094	152.934	160.908	143.928	167.404
Cluster12	111.344	98.453	160.328	141.031	125.100	97.502
Cluster13	153.847	138.216	101.466	168.718	100.478	147.341
Cluster14	0.000	103.215	106.740	147.689	112.843	146.185
Cluster15	103.215	0.000	126.824	110.384	111.043	92.822
Cluster16	106.740	126.824	0.000	179.616	85.828	144.673
Cluster17	147.689	110.384	179.616	0.000	130.809	153.760
Cluster18	112.843	111.043	85.828	130.809	0.000	113.753
Cluster19	146.185	92.822	144.673	153.760	113.753	0.000
Cluster20	158.304	124.272	158.085	132.158	105.496	120.359
	Cluster20					
Cluster1	105.303					
Cluster2	135.688					
Cluster3	185.527					
Cluster4	131.596					
Cluster5	158.050					
Cluster6	167.521					
Cluster7	152.309					
Cluster8	148.101					
Cluster9	124.851					
Cluster10	127.559					
Cluster11	172.857					
Cluster12	150.406					
Cluster13	138.286					
Cluster14	158.304					
Cluster15	124.272					
Cluster16	158.085					
Cluster17	132.158					
Cluster18	105.496					
Cluster19	120.359					
Cluster20	0.000					

Dendrogram



O gráfico e tabela acima mostram a alta similaridade dos dados, tanto que 83% das observações estão concentradas em apenas 2 clusters (cluster 1 com 85 e cluster 2 com 25 obs), portanto não justifica considerar 20 clusters nesta análise.

8.2.2 Análise utilizando somente os PCs (por variáveis e por observações)

Análise com os 4 PCs com 20 clusters (por variáveis e por observações) e Dendrogramas

Cluster Analysis of Observations: PC1, PC2, PC3, PC4

Euclidean Distance, Single Linkage
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	131	98.7552	0.19128	114 115	114	2
2	130	98.4141	0.24369	103 123	103	2
3	129	98.1862	0.27871	93 114	93	3
4	128	97.2262	0.42623	60 110	60	2
5	127	97.1299	0.44103	44 67	44	2
6	126	97.0271	0.45682	4 59	4	2
7	125	96.8351	0.48633	41 60	41	3
8	124	96.7902	0.49322	99 107	99	2
9	123	96.7335	0.50193	2 53	2	2
10	122	96.5366	0.53219	36 38	36	2
11	121	96.2815	0.57138	40 93	40	4
12	120	96.0868	0.60131	32 108	32	2
13	119	96.0476	0.60734	35 36	35	3
14	118	95.9672	0.61968	21 28	21	2
15	117	95.8917	0.63128	35 97	35	4

16	116	95.7692	0.65011	68	72	68	2
17	115	95.6201	0.67302	43	126	43	2
18	114	95.6163	0.67361	27	35	27	5
19	113	95.5936	0.67709	16	95	16	2
20	112	95.5590	0.68241	82	109	82	2
21	111	95.4974	0.69187	13	96	13	2
22	110	95.4384	0.70094	22	40	22	5
23	109	95.3932	0.70789	26	121	26	2
24	108	95.3510	0.71436	6	56	6	2
25	107	95.2380	0.73174	76	98	76	2
26	106	95.0313	0.76349	6	22	6	7
27	105	94.8966	0.78420	8	129	8	2
28	104	94.7924	0.80020	30	106	30	2
29	103	94.7830	0.80165	84	116	84	2
30	102	94.6836	0.81693	33	88	33	2
31	101	94.5041	0.84451	11	127	11	2
32	100	94.4841	0.84758	12	18	12	2
33	99	94.4555	0.85198	50	68	50	3
34	98	94.4372	0.85479	7	43	7	3
35	97	94.3747	0.86439	94	131	94	2
36	96	94.3608	0.86653	26	104	26	3
37	95	94.3421	0.86940	6	7	6	10
38	94	94.2682	0.88075	39	64	39	2
39	93	94.2146	0.88899	6	33	6	12
40	92	94.1221	0.90320	26	84	26	5
41	91	94.0396	0.91588	6	89	6	13
42	90	94.0362	0.91640	75	85	75	2
43	89	94.0224	0.91852	102	124	102	2
44	88	93.9572	0.92854	26	37	26	6
45	87	93.9294	0.93281	5	14	5	2
46	86	93.9125	0.93542	2	111	2	3
47	85	93.8288	0.94828	13	86	13	3
48	84	93.8125	0.95077	30	101	30	3
49	83	93.8012	0.95252	39	99	39	4
50	82	93.7902	0.95421	8	69	8	3
51	81	93.7757	0.95643	3	34	3	2
52	80	93.7153	0.96572	11	41	11	5
53	79	93.6151	0.98112	3	92	3	3
54	78	93.5526	0.99071	81	102	81	3
55	77	93.5380	0.99296	5	42	5	3
56	76	93.5034	0.99828	20	103	20	3
57	75	93.3348	1.02418	32	39	32	6
58	74	93.3290	1.02508	70	113	70	2
59	73	93.3196	1.02652	20	117	20	4
60	72	93.2910	1.03092	8	54	8	4
61	71	93.2750	1.03337	27	130	27	6
62	70	93.2267	1.04080	5	81	5	6
63	69	93.2009	1.04476	6	11	6	18
64	68	93.1637	1.05047	20	74	20	5
65	67	93.0182	1.07284	1	73	1	2
66	66	93.0037	1.07506	5	8	5	10
67	65	92.9176	1.08829	3	87	3	4
68	64	92.8834	1.09355	5	61	5	11
69	63	92.8260	1.10236	12	105	12	3
70	62	92.8178	1.10363	5	62	5	12
71	61	92.8125	1.10444	27	90	27	7
72	60	92.7887	1.10810	48	70	48	3
73	59	92.7442	1.11494	30	82	30	5
74	58	92.6035	1.13655	6	32	6	24
75	57	92.6033	1.13659	6	51	6	25
76	56	92.5251	1.14860	20	44	20	7
77	55	92.5114	1.15070	20	75	20	9
78	54	92.4756	1.15620	10	17	10	2
79	53	92.4256	1.16390	16	27	16	9
80	52	92.3502	1.17547	5	10	5	14
81	51	92.3228	1.17968	5	50	5	17
82	50	92.2246	1.19478	13	16	13	12
83	49	92.1583	1.20496	13	49	13	13
84	48	92.0808	1.21687	1	30	1	7
85	47	92.0455	1.22230	5	66	5	18

86	46	92.0184	1.22646	13	46	13	14
87	45	91.9568	1.23593	2	118	2	4
88	44	91.9346	1.23934	6	25	6	26
89	43	91.8775	1.24811	5	26	5	24
90	42	91.7722	1.26429	6	58	6	27
91	41	91.7204	1.27226	20	21	20	11
92	40	91.6648	1.28080	4	13	4	16
93	39	91.6170	1.28814	77	91	77	2
94	38	91.4558	1.31291	3	9	3	5
95	37	91.4309	1.31674	6	100	6	28
96	36	91.3382	1.33099	24	47	24	2
97	35	91.2670	1.34192	5	122	5	25
98	34	91.2150	1.34991	1	48	1	10
99	33	91.2103	1.35064	2	5	2	29
100	32	91.1761	1.35589	12	20	12	14

Final Partition

Number of clusters: 20

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	60	475.981	2.71447	4.20087
Cluster2	6	11.755	1.34764	1.82405
Cluster3	28	81.312	1.60967	2.81252
Cluster4	17	42.485	1.50234	2.35890
Cluster5	1	0.000	0.00000	0.00000
Cluster6	3	2.319	0.86690	0.97743
Cluster7	1	0.000	0.00000	0.00000
Cluster8	1	0.000	0.00000	0.00000
Cluster9	1	0.000	0.00000	0.00000
Cluster10	1	0.000	0.00000	0.00000
Cluster11	1	0.000	0.00000	0.00000
Cluster12	1	0.000	0.00000	0.00000
Cluster13	2	0.950	0.68905	0.68905
Cluster14	1	0.000	0.00000	0.00000
Cluster15	2	1.208	0.77702	0.77702
Cluster16	2	0.374	0.43220	0.43220
Cluster17	1	0.000	0.00000	0.00000
Cluster18	1	0.000	0.00000	0.00000
Cluster19	1	0.000	0.00000	0.00000
Cluster20	1	0.000	0.00000	0.00000

Cluster Centroids

Variable	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6	Cluster7
PC1	-0.034390	-4.26267	5.16890	-3.96409	0.22869	-6.77750	-8.13369
PC2	0.823283	0.42291	-0.57848	-1.68652	-2.26910	-1.20122	-2.67723
PC3	0.711425	0.06524	-0.79642	-0.93628	-1.28844	-1.19452	-2.01699
PC4	-0.545134	1.09189	0.56372	0.45429	-1.25497	-0.19663	-2.36384

Variable	Cluster8	Cluster9	Cluster10	Cluster11	Cluster12	Cluster13
PC1	3.38342	2.80635	-4.12551	2.98336	-2.89933	1.31972
PC2	-2.46450	1.51104	3.46120	3.55180	-0.39931	-0.06594
PC3	3.56668	0.12352	1.26895	0.39320	1.94931	-1.39402
PC4	-0.47128	0.88649	0.61847	1.62904	0.50840	1.15299

Variable	Cluster14	Cluster15	Cluster16	Cluster17	Cluster18	Cluster19
PC1	-7.00312	1.67211	-5.45267	-6.25648	-5.86159	3.01773
PC2	-1.30180	-0.96676	1.92967	1.25796	-2.34410	-1.32301
PC3	-2.70399	0.47422	1.11733	-0.37244	-1.76740	-0.41917
PC4	0.93113	-0.90993	2.36377	0.89572	-0.28686	-1.82718

Variable	Cluster20	Grand centroid
PC1	-2.58605	-0.000000
PC2	-2.25956	0.000000
PC3	-0.40501	-0.000000
PC4	-1.23919	-0.000000

Distances Between Cluster Centroids

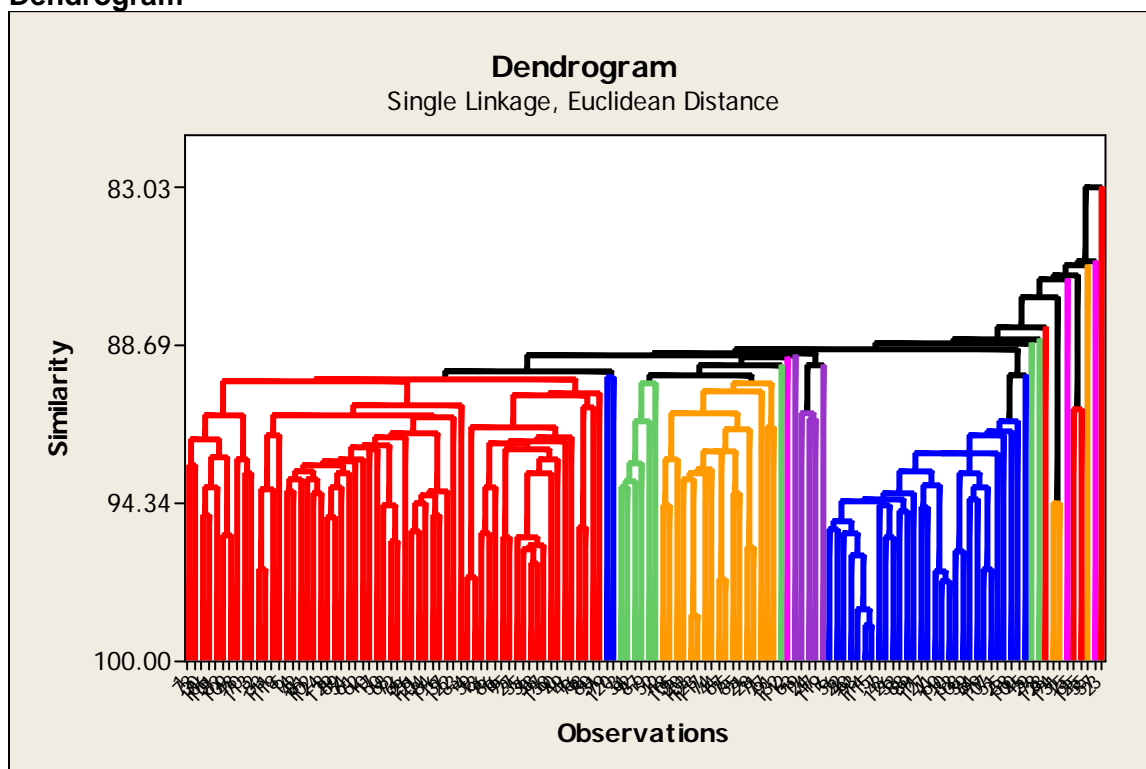
	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6	Cluster7
Cluster1	0.00000	4.59739	5.7046	5.04533	3.75970	7.3022	9.4130
Cluster2	4.59739	0.00000	9.5383	2.43894	5.89571	3.4942	6.3932
Cluster3	5.70459	9.53828	0.0000	9.20167	5.55100	11.9934	13.8356
Cluster4	5.04533	2.43894	9.2017	0.00000	4.57868	2.9396	5.2418
Cluster5	3.75970	5.89571	5.5510	4.57868	0.00000	7.1663	8.4768
Cluster6	7.30221	3.49420	11.9934	2.93959	7.16631	0.0000	3.0645
Cluster7	9.41298	6.39323	13.8356	5.24184	8.47681	3.0645	0.0000
Cluster8	5.53614	9.02793	5.1820	8.70198	5.84610	11.2953	12.9402
Cluster9	3.30726	7.15546	3.3012	7.57451	5.24531	10.1053	12.3439
Cluster10	5.03594	3.30500	10.3428	5.60491	7.86418	5.9585	8.5696
Cluster11	4.62384	7.91776	4.9382	8.88001	7.25372	11.1229	13.5699
Cluster12	3.51353	2.53474	8.5247	3.33470	5.18393	5.1053	7.5212
Cluster13	3.15291	5.79096	3.9728	5.58951	3.44292	8.2895	10.4375
Cluster14	8.18069	4.26371	12.3473	3.56874	7.74710	1.9004	3.8077
Cluster15	2.51112	6.42863	4.0205	6.01126	2.64680	8.6455	10.3634
Cluster16	6.26165	2.53199	11.2255	4.81208	8.29400	4.8433	7.7837
Cluster17	6.49262	2.21419	11.5846	3.79972	7.74352	2.8615	5.6867
Cluster18	7.08519	3.93350	11.2452	2.29630	6.18577	1.5753	3.1064
Cluster19	4.10413	8.05034	3.3227	7.37227	3.12360	9.9610	11.3591
Cluster20	4.21225	3.95750	8.1467	2.31894	2.95018	4.5165	5.9003

	Cluster8	Cluster9	Cluster10	Cluster11	Cluster12	Cluster13
Cluster1	5.5361	3.3073	5.0359	4.6238	3.51353	3.1529
Cluster2	9.0279	7.1555	3.3050	7.9178	2.53474	5.7910
Cluster3	5.1820	3.3012	10.3428	4.9382	8.52470	3.9728
Cluster4	8.7020	7.5745	5.6049	8.8800	3.33470	5.5895
Cluster5	5.8461	5.2453	7.8642	7.2537	5.18393	3.4429
Cluster6	11.2953	10.1053	5.9585	11.1229	5.10530	8.2895
Cluster7	12.9402	12.3439	8.5696	13.5699	7.52122	10.4375
Cluster8	0.0000	5.4623	9.8977	7.1301	6.87849	6.1040
Cluster9	5.4623	0.0000	7.2964	2.1955	6.29926	2.6591
Cluster10	9.8977	7.2964	0.0000	7.2341	4.10877	7.0334
Cluster11	7.1301	2.1955	7.2341	0.0000	7.34129	4.3905
Cluster12	6.8785	6.2993	4.1088	7.3413	0.00000	5.4318
Cluster13	6.1040	2.6591	7.0334	4.3905	5.43184	0.0000
Cluster14	12.2687	10.5894	6.8446	11.5485	6.28391	8.5184
Cluster15	3.8636	3.2827	7.4958	5.3469	5.04059	2.9465
Cluster16	10.5557	8.4591	2.6788	8.6523	4.00977	7.5908
Cluster17	11.1432	9.0799	3.4880	9.5792	4.42240	7.7628
Cluster18	10.6757	9.7441	6.8377	11.0152	5.19650	7.6794
Cluster19	4.3775	3.9667	9.0965	6.0308	6.85050	3.7809
Cluster20	7.2139	6.9349	6.4304	8.5822	3.48649	5.1738

	Cluster14	Cluster15	Cluster16	Cluster17	Cluster18	Cluster19
Cluster1	8.1807	2.5111	6.2617	6.4926	7.0852	4.1041
Cluster2	4.2637	6.4286	2.5320	2.2142	3.9335	8.0503
Cluster3	12.3473	4.0205	11.2255	11.5846	11.2452	3.3227
Cluster4	3.5687	6.0113	4.8121	3.7997	2.2963	7.3723
Cluster5	7.7471	2.6468	8.2940	7.7435	6.1858	3.1236
Cluster6	1.9004	8.6455	4.8433	2.8615	1.5753	9.9610
Cluster7	3.8077	10.3634	7.7837	5.6867	3.1064	11.3591
Cluster8	12.2687	3.8636	10.5557	11.1432	10.6757	4.3775
Cluster9	10.5894	3.2827	8.4591	9.0799	9.7441	3.9667
Cluster10	6.8446	7.4958	2.6788	3.4880	6.8377	9.0965
Cluster11	11.5485	5.3469	8.6523	9.5792	11.0152	6.0308
Cluster12	6.2839	5.0406	4.0098	4.4224	5.1965	6.8505
Cluster13	8.5184	2.9465	7.5908	7.7628	7.6794	3.7809
Cluster14	0.0000	9.4267	5.4315	3.5422	2.1795	10.6417
Cluster15	9.4267	0.0000	8.3835	8.4728	8.0042	1.8913
Cluster16	5.4315	8.3835	0.0000	2.3392	5.8120	10.1120
Cluster17	3.5422	8.4728	2.3392	0.0000	4.0590	10.0044
Cluster18	2.1795	8.0042	5.8120	4.0590	0.0000	9.1693
Cluster19	10.6417	1.8913	10.1120	10.0044	9.1693	0.0000
Cluster20	5.5157	4.5480	6.4083	5.5140	3.6742	5.7119

	Cluster20
Cluster1	4.21225
Cluster2	3.95750
Cluster3	8.14672
Cluster4	2.31894
Cluster5	2.95018
Cluster6	4.51649
Cluster7	5.90034
Cluster8	7.21391
Cluster9	6.93493
Cluster10	6.43040
Cluster11	8.58220
Cluster12	3.48649
Cluster13	5.17375
Cluster14	5.51575
Cluster15	4.54804
Cluster16	6.40827
Cluster17	5.51398
Cluster18	3.67415
Cluster19	5.71186
Cluster20	0.00000

Dendrogram



Neste caso, mesmo considerando apenas os 4 PCs, os resultados foram muito parecidos com o anterior, com todas as variáveis, onde 84% das observações se apresentaram concentradas praticamente nos 4 primeiros clusters, portanto este modelo é igualmente insuficiente para explicar os dados.

8.2.3 Análise utilizando somente os PCs (por variáveis e por observações)

Análise com os 4 PCs com 30 clusters (por variáveis e por observações) e Dendrogramas

Cluster Analysis of Observations: PC1, PC2, PC3, PC4

Euclidean Distance, Single Linkage
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	131	98.7552	0.19128	114 115	114	2
2	130	98.4141	0.24369	103 123	103	2
3	129	98.1862	0.27871	93 114	93	3
4	128	97.2262	0.42623	60 110	60	2
5	127	97.1299	0.44103	44 67	44	2
6	126	97.0271	0.45682	4 59	4	2
7	125	96.8351	0.48633	41 60	41	3
8	124	96.7902	0.49322	99 107	99	2
9	123	96.7335	0.50193	2 53	2	2
10	122	96.5366	0.53219	36 38	36	2
11	121	96.2815	0.57138	40 93	40	4
12	120	96.0868	0.60131	32 108	32	2
13	119	96.0476	0.60734	35 36	35	3
14	118	95.9672	0.61968	21 28	21	2
15	117	95.8917	0.63128	35 97	35	4
16	116	95.7692	0.65011	68 72	68	2
17	115	95.6201	0.67302	43 126	43	2
18	114	95.6163	0.67361	27 35	27	5
19	113	95.5936	0.67709	16 95	16	2
20	112	95.5590	0.68241	82 109	82	2
21	111	95.4974	0.69187	13 96	13	2
22	110	95.4384	0.70094	22 40	22	5
23	109	95.3932	0.70789	26 121	26	2
24	108	95.3510	0.71436	6 56	6	2
25	107	95.2380	0.73174	76 98	76	2
26	106	95.0313	0.76349	6 22	6	7
27	105	94.8966	0.78420	8 129	8	2
28	104	94.7924	0.80020	30 106	30	2
29	103	94.7830	0.80165	84 116	84	2
30	102	94.6836	0.81693	33 88	33	2
31	101	94.5041	0.84451	11 127	11	2
32	100	94.4841	0.84758	12 18	12	2
33	99	94.4555	0.85198	50 68	50	3
34	98	94.4372	0.85479	7 43	7	3
35	97	94.3747	0.86439	94 131	94	2
36	96	94.3608	0.86653	26 104	26	3
37	95	94.3421	0.86940	6 7	6	10
38	94	94.2682	0.88075	39 64	39	2
39	93	94.2146	0.88899	6 33	6	12
40	92	94.1221	0.90320	26 84	26	5
41	91	94.0396	0.91588	6 89	6	13
42	90	94.0362	0.91640	75 85	75	2
43	89	94.0224	0.91852	102 124	102	2
44	88	93.9572	0.92854	26 37	26	6
45	87	93.9294	0.93281	5 14	5	2
46	86	93.9125	0.93542	2 111	2	3
47	85	93.8288	0.94828	13 86	13	3
48	84	93.8125	0.95077	30 101	30	3
49	83	93.8012	0.95252	39 99	39	4
50	82	93.7902	0.95421	8 69	8	3
51	81	93.7757	0.95643	3 34	3	2
52	80	93.7153	0.96572	11 41	11	5
53	79	93.6151	0.98112	3 92	3	3
54	78	93.5526	0.99071	81 102	81	3
55	77	93.5380	0.99296	5 42	5	3
56	76	93.5034	0.99828	20 103	20	3
57	75	93.3348	1.02418	32 39	32	6
58	74	93.3290	1.02508	70 113	70	2
59	73	93.3196	1.02652	20 117	20	4
60	72	93.2910	1.03092	8 54	8	4
61	71	93.2750	1.03337	27 130	27	6

62	70	93.2267	1.04080	5	81	5	6
63	69	93.2009	1.04476	6	11	6	18
64	68	93.1637	1.05047	20	74	20	5
65	67	93.0182	1.07284	1	73	1	2
66	66	93.0037	1.07506	5	8	5	10
67	65	92.9176	1.08829	3	87	3	4
68	64	92.8834	1.09355	5	61	5	11
69	63	92.8260	1.10236	12	105	12	3
70	62	92.8178	1.10363	5	62	5	12
71	61	92.8125	1.10444	27	90	27	7
72	60	92.7887	1.10810	48	70	48	3
73	59	92.7442	1.11494	30	82	30	5
74	58	92.6035	1.13655	6	32	6	24
75	57	92.6033	1.13659	6	51	6	25
76	56	92.5251	1.14860	20	44	20	7
77	55	92.5114	1.15070	20	75	20	9
78	54	92.4756	1.15620	10	17	10	2
79	53	92.4256	1.16390	16	27	16	9
80	52	92.3502	1.17547	5	10	5	14
81	51	92.3228	1.17968	5	50	5	17
82	50	92.2246	1.19478	13	16	13	12
83	49	92.1583	1.20496	13	49	13	13
84	48	92.0808	1.21687	1	30	1	7
85	47	92.0455	1.22230	5	66	5	18
86	46	92.0184	1.22646	13	46	13	14
87	45	91.9568	1.23593	2	118	2	4
88	44	91.9346	1.23934	6	25	6	26
89	43	91.8775	1.24811	5	26	5	24
90	42	91.7722	1.26429	6	58	6	27
91	41	91.7204	1.27226	20	21	20	11
92	40	91.6648	1.28080	4	13	4	16
93	39	91.6170	1.28814	77	91	77	2
94	38	91.4558	1.31291	3	9	3	5
95	37	91.4309	1.31674	6	100	6	28
96	36	91.3382	1.33099	24	47	24	2
97	35	91.2670	1.34192	5	122	5	25
98	34	91.2150	1.34991	1	48	1	10
99	33	91.2103	1.35064	2	5	2	29
100	32	91.1761	1.35589	12	20	12	14

Final Partition

Number of clusters: 30

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	10	22.959	1.41578	2.58997
Cluster2	29	158.006	2.17945	4.05434
Cluster3	5	7.763	1.17763	1.70476
Cluster4	16	31.694	1.30496	2.34332
Cluster5	28	81.312	1.60967	2.81252
Cluster6	14	28.034	1.35581	1.85350
Cluster7	1	0.000	0.00000	0.00000
Cluster8	3	2.319	0.86690	0.97743
Cluster9	1	0.000	0.00000	0.00000
Cluster10	1	0.000	0.00000	0.00000
Cluster11	1	0.000	0.00000	0.00000
Cluster12	1	0.000	0.00000	0.00000
Cluster13	1	0.000	0.00000	0.00000
Cluster14	1	0.000	0.00000	0.00000
Cluster15	1	0.000	0.00000	0.00000
Cluster16	1	0.000	0.00000	0.00000
Cluster17	2	0.950	0.68905	0.68905
Cluster18	1	0.000	0.00000	0.00000
Cluster19	2	0.268	0.36587	0.36587
Cluster20	2	0.830	0.64407	0.64407
Cluster21	1	0.000	0.00000	0.00000
Cluster22	1	0.000	0.00000	0.00000

Cluster23	1	0.000	0.00000	0.00000
Cluster24	1	0.000	0.00000	0.00000
Cluster25	2	0.374	0.43220	0.43220
Cluster26	1	0.000	0.00000	0.00000
Cluster27	1	0.000	0.00000	0.00000
Cluster28	1	0.000	0.00000	0.00000
Cluster29	1	0.000	0.00000	0.00000
Cluster30	1	0.000	0.00000	0.00000

Cluster Centroids

Variable	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6	Cluster7
PC1	-0.17771	-0.36273	-4.41708	0.43718	5.16890	-3.71355	0.22869
PC2	0.45663	2.24372	0.14422	-1.49870	-0.57848	-1.65660	-2.26910
PC3	0.01435	-0.09745	-0.05626	2.57561	-0.79642	-0.78641	-1.28844
PC4	-1.44337	-0.24755	0.96224	-0.68371	0.56372	0.20966	-1.25497

Variable	Cluster8	Cluster9	Cluster10	Cluster11	Cluster12	Cluster13
PC1	-6.77750	-8.13369	3.38342	0.53971	2.80635	-3.49060
PC2	-1.20122	-2.67723	-2.46450	1.20791	1.51104	1.81639
PC3	-1.19452	-2.01699	3.56668	1.80326	0.12352	0.67273
PC4	-0.19663	-2.36384	-0.47128	-0.73527	0.88649	1.74014

Variable	Cluster14	Cluster15	Cluster16	Cluster17	Cluster18	Cluster19
PC1	-4.12551	2.98336	-2.89933	1.31972	-4.88770	0.96182
PC2	3.46120	3.55180	-0.39931	-0.06594	-2.03148	0.24482
PC3	1.26895	0.39320	1.94931	-1.39402	-0.53012	0.75147
PC4	0.61847	1.62904	0.50840	1.15299	1.40659	1.09514

Variable	Cluster20	Cluster21	Cluster22	Cluster23	Cluster24	Cluster25
PC1	-5.25606	-7.00312	2.00160	0.86966	-0.09497	-5.45267
PC2	-1.72348	-1.30180	-0.73249	0.66688	1.37770	1.92967
PC3	-2.18845	-2.70399	0.30798	1.94412	-1.09186	1.11733
PC4	1.69054	0.93113	-0.26753	0.58363	-2.19451	2.36377

Variable	Cluster26	Cluster27	Cluster28	Cluster29	Cluster30	Grand centroid
PC1	-6.25648	-5.86159	1.34261	3.01773	-2.58605	-0.000000
PC2	1.25796	-2.34410	-1.20103	-1.32301	-2.25956	0.000000
PC3	-0.37244	-1.76740	0.64046	-0.41917	-0.40501	-0.000000
PC4	0.89572	-0.28686	-1.55233	-1.82718	-1.23919	-0.000000

Distances Between Cluster Centroids

	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6	Cluster7
Cluster1	0.00000	2.16111	4.88486	3.36727	5.8603	4.51017	3.05409
Cluster2	2.16111	0.00000	4.72345	4.68839	6.3016	5.20808	4.81130
Cluster3	4.88486	4.72345	0.00000	5.99157	9.6499	2.19941	5.81735
Cluster4	3.36727	4.68839	5.99157	0.00000	6.0135	5.41800	3.98675
Cluster5	5.86033	6.30163	9.64988	6.01354	0.0000	8.95465	5.55100
Cluster6	4.51017	5.20808	2.19941	5.41800	8.9546	0.00000	4.27943
Cluster7	3.05409	4.81130	5.81735	3.98675	5.5510	4.27943	0.00000
Cluster8	7.02293	7.36363	3.16550	8.16034	11.9934	3.15069	7.16631
Cluster9	8.83698	9.63155	6.05648	9.93798	13.8356	5.35878	8.47681
Cluster10	5.89734	7.04818	9.10131	3.26197	5.1820	8.39242	5.84610
Cluster11	2.18649	2.39538	5.66044	2.81700	5.7503	5.82193	4.69202
Cluster12	3.93148	3.45178	7.35420	4.81142	3.3012	7.33683	5.24531
Cluster13	4.83657	3.80923	2.18885	5.99277	9.1796	4.07220	6.58340
Cluster14	5.51705	4.27292	3.60024	6.98726	10.3428	5.54556	7.86418
Cluster15	5.39960	4.08286	8.18687	6.48863	4.9382	8.68228	7.25372
Cluster16	3.96148	4.26387	2.61291	3.76222	8.5247	3.13324	5.18393
Cluster17	3.35260	3.43624	5.89755	4.68648	3.9728	5.39658	3.44292
Cluster18	6.06576	6.45570	2.31887	6.53093	10.1994	1.73710	5.82176
Cluster19	2.88630	2.87637	5.44177	3.13160	4.5887	5.35011	4.06713
Cluster20	6.72431	6.91461	3.04448	7.79723	10.6395	2.55784	6.31397
Cluster21	7.91871	8.05288	3.97367	9.26708	12.3473	3.89164	7.74710
Cluster22	2.76265	3.82265	6.60402	2.88960	3.4592	5.91120	3.00466
Cluster23	2.99565	2.97727	5.68927	2.62330	5.2484	5.83095	4.78129
Cluster24	1.62577	2.36668	5.58924	4.92846	6.2634	5.30795	3.78487

Cluster25	6.76061	5.85668	2.75714	7.60643	11.2255	4.91429	8.29400
Cluster26	6.57378	6.09021	2.17444	7.97432	11.5846	3.95009	7.74352
Cluster27	6.68302	7.35363	3.57304	7.70769	11.2452	2.50911	6.18577
Cluster28	2.33732	4.12570	6.46469	2.32545	4.6444	5.55993	2.48810
Cluster29	3.70312	5.17182	8.08342	4.11902	3.3227	7.05017	3.12360
Cluster30	3.65996	5.12841	3.75483	4.34873	8.1467	1.96963	2.95018

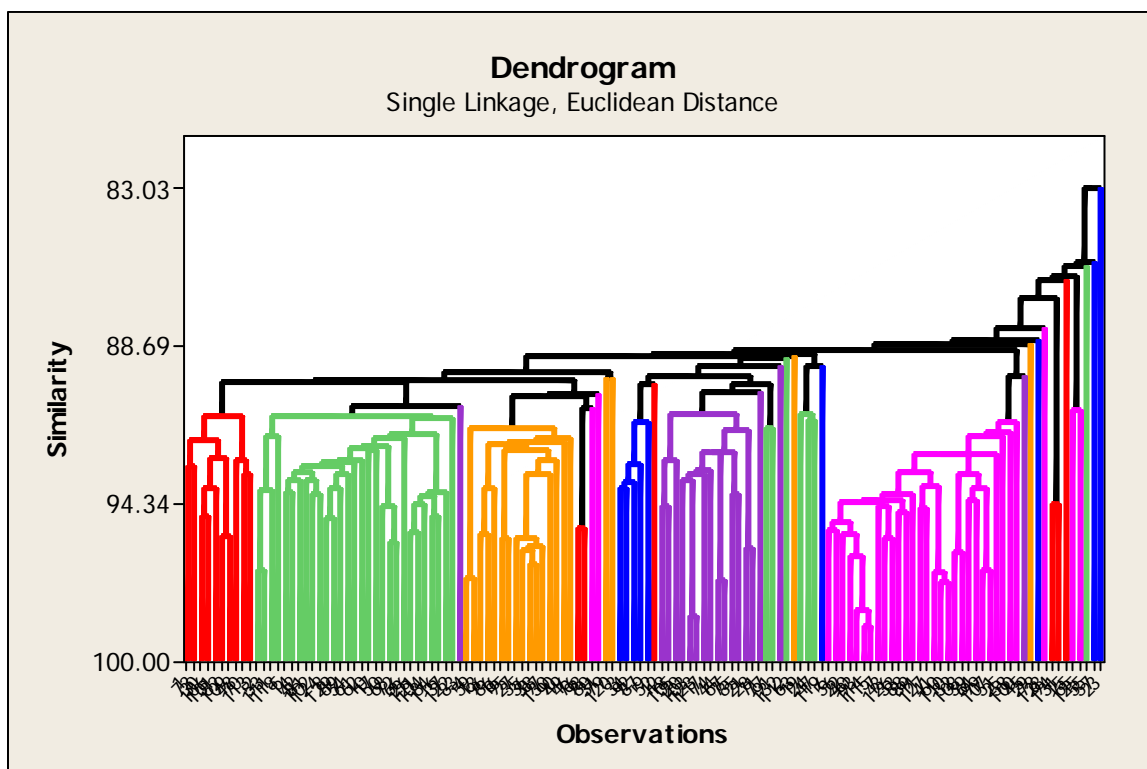
	Cluster8	Cluster9	Cluster10	Cluster11	Cluster12	Cluster13
Cluster1	7.0229	8.8370	5.8973	2.1865	3.9315	4.83657
Cluster2	7.3636	9.6315	7.0482	2.3954	3.4518	3.80923
Cluster3	3.1655	6.0565	9.1013	5.6604	7.3542	2.18885
Cluster4	8.1603	9.9380	3.2620	2.8170	4.8114	5.99277
Cluster5	11.9934	13.8356	5.1820	5.7503	3.3012	9.17959
Cluster6	3.1507	5.3588	8.3924	5.8219	7.3368	4.07220
Cluster7	7.1663	8.4768	5.8461	4.6920	5.2453	6.58340
Cluster8	0.0000	3.0645	11.2953	8.2839	10.1053	5.21031
Cluster9	3.0645	0.0000	12.9402	10.3715	12.3439	8.11345
Cluster10	11.2953	12.9402	0.0000	4.9752	5.4623	8.87939
Cluster11	8.2839	10.3715	4.9752	0.0000	3.2682	4.90096
Cluster12	10.1053	12.3439	5.4623	3.2682	0.0000	6.38554
Cluster13	5.2103	8.1135	8.8794	4.9010	6.3855	0.00000
Cluster14	5.9585	8.5696	9.8977	5.3814	7.2964	2.17305
Cluster15	11.1229	13.5699	7.1301	4.3639	2.1955	6.70927
Cluster16	5.1053	7.5212	6.8785	3.9973	6.2993	2.89927
Cluster17	8.2895	10.4375	6.1040	4.0024	2.6591	5.59451
Cluster18	2.6967	5.2326	9.4292	7.0698	8.5115	4.27973
Cluster19	8.2124	10.5307	4.8563	2.3585	2.3331	4.76614
Cluster20	2.6715	5.0654	10.6294	8.0001	9.0253	4.88225
Cluster21	1.9004	3.8077	12.2687	9.2890	10.5894	5.84103
Cluster22	8.9194	10.7845	3.9459	2.8908	2.6546	6.38943
Cluster23	8.5105	10.7991	4.4576	1.4700	2.8053	4.82569
Cluster24	7.4370	9.0525	7.1787	3.3080	4.4051	5.50618
Cluster25	4.8433	7.7837	10.5557	6.8194	8.4591	2.10929
Cluster26	2.8615	5.6867	11.1432	7.3201	9.0799	3.12528
Cluster27	1.5753	3.1064	10.6757	8.1575	9.7441	5.74406
Cluster28	8.4345	9.9850	3.9361	2.9099	3.9640	6.58074
Cluster29	9.9610	11.3591	4.3775	4.3218	3.9667	8.13217
Cluster30	4.5165	5.9003	7.2139	5.1888	6.9349	5.24114

	Cluster14	Cluster15	Cluster16	Cluster17	Cluster18	Cluster19
Cluster1	5.5170	5.3996	3.96148	3.3526	6.0658	2.8863
Cluster2	4.2729	4.0829	4.26387	3.4362	6.4557	2.8764
Cluster3	3.6002	8.1869	2.61291	5.8976	2.3189	5.4418
Cluster4	6.9873	6.4886	3.76222	4.6865	6.5309	3.1316
Cluster5	10.3428	4.9382	8.52470	3.9728	10.1994	4.5887
Cluster6	5.5456	8.6823	3.13324	5.3966	1.7371	5.3501
Cluster7	7.8642	7.2537	5.18393	3.4429	5.8218	4.0671
Cluster8	5.9585	11.1229	5.10530	8.2895	2.6967	8.2124
Cluster9	8.5696	13.5699	7.52122	10.4375	5.2326	10.5307
Cluster10	9.8977	7.1301	6.87849	6.1040	9.4292	4.8563
Cluster11	5.3814	4.3639	3.99728	4.0024	7.0698	2.3585
Cluster12	7.2964	2.1955	6.29926	2.6591	8.5115	2.3331
Cluster13	2.1730	6.7093	2.89927	5.5945	4.2797	4.7661
Cluster14	0.0000	7.2341	4.10877	7.0334	5.8829	6.0598
Cluster15	7.2341	0.0000	7.34129	4.3905	9.6968	3.9289
Cluster16	4.1088	7.3413	0.00000	5.4318	3.6840	4.1355
Cluster17	7.0334	4.3905	5.43184	0.0000	6.5731	2.1980
Cluster18	5.8829	9.6968	3.68401	6.5731	0.0000	6.4139
Cluster19	6.0598	3.9289	4.13551	2.1980	6.4139	0.0000
Cluster20	6.4235	10.1186	5.08194	6.8490	1.7496	7.1787
Cluster21	6.8446	11.5485	6.28391	8.5184	3.1558	8.8204
Cluster22	7.5390	4.7878	5.23700	2.4133	7.2564	2.0223
Cluster23	5.7634	4.0359	3.91762	3.4938	6.8722	1.3677
Cluster24	5.8372	5.5703	5.25134	3.9221	6.9193	4.0767
Cluster25	2.6788	8.6523	4.00977	7.5908	4.4317	6.7622
Cluster26	3.4880	9.5792	4.42240	7.7628	3.6028	7.3779
Cluster27	6.8377	11.0152	5.19650	7.6794	2.3334	7.8432
Cluster28	7.5329	5.9551	4.95949	3.5703	7.0450	3.0425
Cluster29	9.0965	6.0308	6.85050	3.7809	8.5713	4.0737
Cluster30	6.4304	8.5822	3.48649	5.1738	3.5165	5.0642

	Cluster20	Cluster21	Cluster22	Cluster23	Cluster24	Cluster25
Cluster1	6.7243	7.9187	2.7626	2.9957	1.62577	6.7606
Cluster2	6.9146	8.0529	3.8227	2.9773	2.36668	5.8567
Cluster3	3.0445	3.9737	6.6040	5.6893	5.58924	2.7571
Cluster4	7.7972	9.2671	2.8896	2.6233	4.92846	7.6064
Cluster5	10.6395	12.3473	3.4592	5.2484	6.26339	11.2255
Cluster6	2.5578	3.8916	5.9112	5.8310	5.30795	4.9143
Cluster7	6.3140	7.7471	3.0047	4.7813	3.78487	8.2940
Cluster8	2.6715	1.9004	8.9194	8.5105	7.43701	4.8433
Cluster9	5.0654	3.8077	10.7845	10.7991	9.05250	7.7837
Cluster10	10.6294	12.2687	3.9459	4.4576	7.17866	10.5557
Cluster11	8.0001	9.2890	2.8908	1.4700	3.30798	6.8194
Cluster12	9.0253	10.5894	2.6546	2.8053	4.40513	8.4591
Cluster13	4.8822	5.8410	6.3894	4.8257	5.50618	2.1093
Cluster14	6.4235	6.8446	7.5390	5.7634	5.83717	2.6788
Cluster15	10.1186	11.5485	4.7878	4.0359	5.57026	8.6523
Cluster16	5.0819	6.2839	5.2370	3.9176	5.25134	4.0098
Cluster17	6.8490	8.5184	2.4133	3.4938	3.92205	7.5908
Cluster18	1.7496	3.1558	7.2564	6.8722	6.91926	4.4317
Cluster19	7.1787	8.8204	2.0223	1.3677	4.07675	6.7622
Cluster20	0.0000	2.0181	7.9826	7.8449	7.24916	4.9765
Cluster21	2.0181	0.0000	9.5874	9.3585	8.20188	5.4315
Cluster22	7.9826	9.5874	0.0000	2.5770	3.81069	8.3805
Cluster23	7.8449	9.3585	2.5770	0.0000	4.28614	6.7394
Cluster24	7.2492	8.2019	3.8107	4.2861	0.00000	7.3938
Cluster25	4.9765	5.4315	8.3805	6.7394	7.39378	0.0000
Cluster26	3.7175	3.5422	8.6008	7.5230	6.93149	2.3392
Cluster27	2.1998	2.1795	8.2906	8.3011	7.15550	5.8120
Cluster28	7.8952	9.3281	1.5540	3.1583	3.48278	8.4581
Cluster29	9.1717	10.6417	2.0839	4.4689	4.19167	10.1120
Cluster30	4.3795	5.5157	4.9831	5.4173	4.56284	6.4083

	Cluster26	Cluster27	Cluster28	Cluster29	Cluster30
Cluster1	6.5738	6.6830	2.33732	3.7031	3.65996
Cluster2	6.0902	7.3536	4.12570	5.1718	5.12841
Cluster3	2.1744	3.5730	6.46469	8.0834	3.75483
Cluster4	7.9743	7.7077	2.32545	4.1190	4.34873
Cluster5	11.5846	11.2452	4.64439	3.3227	8.14672
Cluster6	3.9501	2.5091	5.55993	7.0502	1.96963
Cluster7	7.7435	6.1858	2.48810	3.1236	2.95018
Cluster8	2.8615	1.5753	8.43453	9.9610	4.51649
Cluster9	5.6867	3.1064	9.98498	11.3591	5.90034
Cluster10	11.1432	10.6757	3.93608	4.3775	7.21391
Cluster11	7.3201	8.1575	2.90987	4.3218	5.18885
Cluster12	9.0799	9.7441	3.96395	3.9667	6.93493
Cluster13	3.1253	5.7441	6.58074	8.1322	5.24114
Cluster14	3.4880	6.8377	7.53287	9.0965	6.43040
Cluster15	9.5792	11.0152	5.95514	6.0308	8.58220
Cluster16	4.4224	5.1965	4.95949	6.8505	3.48649
Cluster17	7.7628	7.6794	3.57027	3.7809	5.17375
Cluster18	3.6028	2.3334	7.04499	8.5713	3.51645
Cluster19	7.3779	7.8432	3.04252	4.0737	5.06418
Cluster20	3.7175	2.1998	7.89521	9.1717	4.37954
Cluster21	3.5422	2.1795	9.32815	10.6417	5.51575
Cluster22	8.6008	8.2906	1.55404	2.0839	4.98307
Cluster23	7.5230	8.3011	3.15826	4.4689	5.41729
Cluster24	6.9315	7.1555	3.48278	4.1917	4.56284
Cluster25	2.3392	5.8120	8.45815	10.1120	6.40827
Cluster26	0.0000	4.0590	8.41497	10.0044	5.51398
Cluster27	4.0590	0.0000	7.78501	9.1693	3.67415
Cluster28	8.4150	7.7850	0.00000	2.0048	4.21259
Cluster29	10.0044	9.1693	2.00480	0.0000	5.71186
Cluster30	5.5140	3.6742	4.21259	5.7119	0.00000

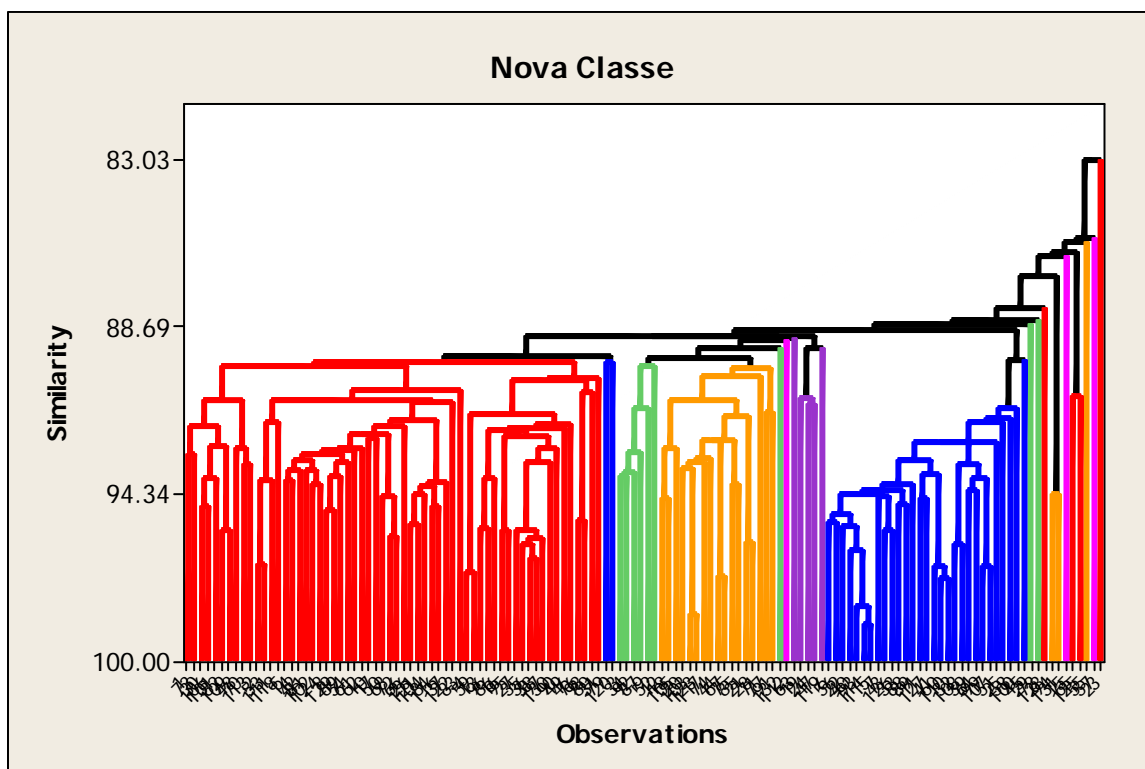
Dendrogram



O teste com 30 clusters também não se mostrou eficiente, 72% das observações ficaram concentradas em apenas 6 clusters.

8.2.4 Categorização (all regions)

Nesta fase será criada uma nova coluna na base de dados para comportar os membros do cluster.



O gráfico acima reflete um bom grau de similaridade entre as observações. Apesar de destacar 3 grandes blocos (vermelho, laranja e azul), vemos que as observações em média estão acima de 90% de similaridade.

8.2.5 Categorização (regions: AIBER e AVECO)

A partir daqui, os dados estão ordenados por bloco e separados para análise. Somente as regiões AIBER e AVECO constarão das análises. A região OTHERS, foi eliminada.

Realizado testes de similaridade com AIBER e AVECO, utilizando 1, 10, 12, 20 e 30 clusters, dos quais o de 12 clusters, aparentemente, apresentou o melhor grau de similaridade e será utilizado nas análises das próximas sessões.

8.2.6 Análise AIBER com 12 clusters (por variáveis e por observações)

Análise AIBER com 12 clusters (por variáveis e por observações) e Dendrogramas

Cluster Analysis of Variables: EPI-2014_n_1, Governança_n, Biodiversity, ...

Correlation Coefficient Distance, Single Linkage
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	18	98.0980	0.038040	15 19	15	2
2	17	97.7194	0.045613	5 15	5	3

3	16	95.9319	0.081361	4	5	4	4
4	15	94.7694	0.104612	2	4	2	5
5	14	93.8810	0.122379	2	6	2	6
6	13	91.5211	0.169577	1	2	1	7
7	12	90.6321	0.187359	10	17	10	2
8	11	81.3907	0.372187	1	7	1	8
9	10	78.3403	0.433194	8	16	8	2
10	9	75.2882	0.494235	1	11	1	9
11	8	75.0539	0.498921	13	18	13	2
12	7	74.8844	0.502312	1	8	1	11
13	6	71.5895	0.568210	1	13	1	13
14	5	70.3565	0.592871	3	10	3	3
15	4	70.0104	0.599791	1	3	1	16
16	3	69.7355	0.605290	1	12	1	17
17	2	66.6408	0.667184	1	9	1	18
18	1	55.4307	0.891386	1	14	1	19

Final Partition

Cluster 1

EPI-2014_n_1 Governança_n_1 Democracy Index-2014_n_1 Opportunity_n_1
IDH-2014_n_1 PC1_1 SPI-2014_n_1

Cluster 2

Biodiversity and habitat_n_1

Cluster 3

Tolerance for immigrants_n_1

Cluster 4

EV-Forests_n_1

Cluster 5

Water withdrawals_np_1

Cluster 6

HPI_n_1 PC3_1

Cluster 7

Women treated with respect_n_1

Cluster 8

EV-Fisheries_np_1

Cluster 9

Health and Wellness_n_1

Cluster 10

Religious tolerance_n_1

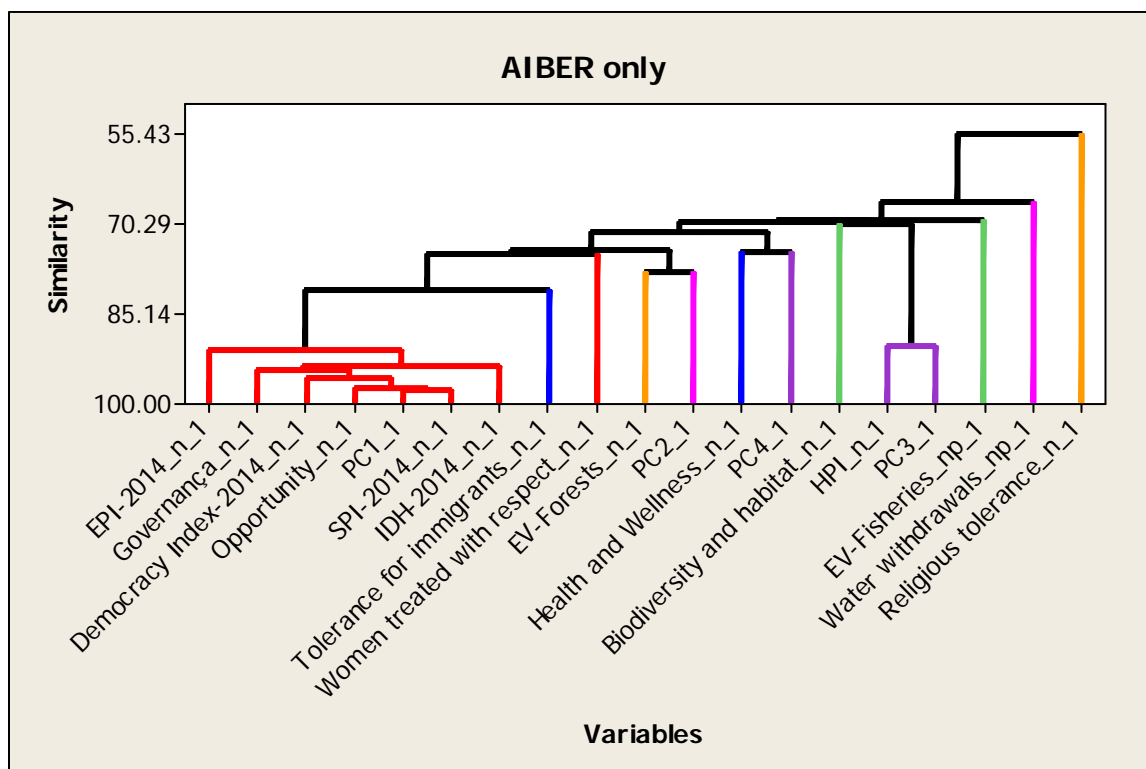
Cluster 11

PC2_1

Cluster 12

PC4_1

Dendrogram



No gráfico e tabela acima, notamos que mais de 50% das observações estão concentradas nos primeiros 4 clusters. O cluster 1 é a partição que abriga a maior quantidade de variáveis nesta análise, são elas: EPI-2014, Governança, Democracy Index-2014, Opportunity, IDH-2014, PC1 e SPI-2014.

Cluster Analysis of Observations: EPI-2014_n_1, Governança_n, Biodiversidade, ...

Euclidean Distance, Single Linkage
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	20	79.3730	31.3986	12 14	12	2
2	19	76.6704	35.5125	11 12	11	3
3	18	71.6977	43.0820	2 11	2	4
4	17	70.3577	45.1218	2 9	2	5
5	16	70.0311	45.6189	2 17	2	6
6	15	70.0169	45.6405	2 8	2	7
7	14	68.0779	48.5921	2 10	2	8
8	13	67.4285	49.5807	1 3	1	2
9	12	65.7116	52.1941	1 5	1	3
10	11	65.2982	52.8234	2 15	2	9
11	10	63.8306	55.0574	1 2	1	12
12	9	62.0624	57.7490	1 6	1	13
13	8	61.3253	58.8710	18 19	18	2
14	7	58.7263	62.8272	1 16	1	14
15	6	58.4844	63.1955	1 21	1	15
16	5	50.1223	75.9243	1 13	1	16
17	4	46.6872	81.1532	4 18	4	3
18	3	46.2333	81.8442	4 20	4	4
19	2	45.1611	83.4763	1 4	1	20
20	1	36.8882	96.0693	1 7	1	21

Final Partition

Number of clusters: 12

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	3	4165.0	35.2947	44.2407
Cluster2	8	13691.7	40.7214	52.5045
Cluster3	1	0.0	0.0000	0.0000
Cluster4	1	0.0	0.0000	0.0000
Cluster5	1	0.0	0.0000	0.0000
Cluster6	1	0.0	0.0000	0.0000
Cluster7	1	0.0	0.0000	0.0000
Cluster8	1	0.0	0.0000	0.0000
Cluster9	1	0.0	0.0000	0.0000
Cluster10	1	0.0	0.0000	0.0000
Cluster11	1	0.0	0.0000	0.0000
Cluster12	1	0.0	0.0000	0.0000

Cluster Centroids

Variable	Cluster1	Cluster2	Cluster3	Cluster4
EPI-2014_n_1	76.7513	60.816	75.719	80.123
Governança_n_1	39.5052	32.756	80.674	64.569
Biodiversity and habitat_n_1	64.0067	72.874	60.160	89.550
Democracy Index-2014_n_1	64.3760	53.984	74.763	77.488
Opportunity_n_1	57.5249	45.962	74.961	71.790
IDH-2014_n_1	70.8654	54.296	81.247	70.093
Tolerance for immigrants_n_1	68.0365	54.623	65.753	78.082
EV-Forests_n_1	12.3500	11.788	100.000	19.870
Water withdrawals_np_1	74.9674	67.975	35.893	61.173
HPI_n_1	79.5956	74.538	75.503	100.000
Women treated with respect_n_1	8.1301	16.616	24.390	24.390
EV-Fisheries_np_1	61.1191	51.514	62.598	39.917
Health and Wellness_n_1	76.0376	82.741	74.437	95.350
Religious tolerance_n_1	55.5556	100.000	66.667	100.000
PC1_1	1.3179	-0.078	3.546	3.383
PC2_1	-1.2851	-1.598	-0.369	-2.465
PC3_1	2.8997	2.886	1.004	3.567
PC4_1	-0.7688	-0.572	-0.848	-0.471
SPI-2014_n_1	65.8998	55.996	78.541	81.147

Variable	Cluster5	Cluster6	Cluster7	Cluster8
EPI-2014_n_1	78.275	68.1038	76.334	62.0754
Governança_n_1	28.758	41.7995	47.875	27.3252
Biodiversity and habitat_n_1	56.060	62.3200	86.960	31.2000
Democracy Index-2014_n_1	24.052	61.4929	66.232	56.5166
Opportunity_n_1	29.325	52.3705	59.922	53.1397
IDH-2014_n_1	70.641	68.4919	72.432	55.5570
Tolerance for immigrants_n_1	41.096	30.1370	45.205	91.7808
EV-Forests_n_1	100.000	19.8700	14.350	0.0000
Water withdrawals_np_1	36.156	29.6789	91.560	99.7770
HPI_n_1	81.060	73.1163	84.950	56.0612
Women treated with respect_n_1	54.354	41.4634	36.585	18.2927
EV-Fisheries_np_1	47.589	59.4105	55.550	71.0235
Health and Wellness_n_1	71.737	75.4575	91.059	85.2985
Religious tolerance_n_1	100.000	33.3333	100.000	66.6667
PC1_1	0.540	0.8697	2.056	-0.1236
PC2_1	1.208	0.6669	-1.852	-1.9864
PC3_1	1.803	1.9441	2.611	2.0242
PC4_1	-0.735	0.5836	-0.600	-0.8218
SPI-2014_n_1	51.168	60.7656	71.855	54.0079

Variable	Cluster9	Cluster10	Cluster11	Cluster12
EPI-2014_n_1	96.174	96.6965	68.832	73.0310
Governança_n_1	74.026	70.1330	68.430	8.0062

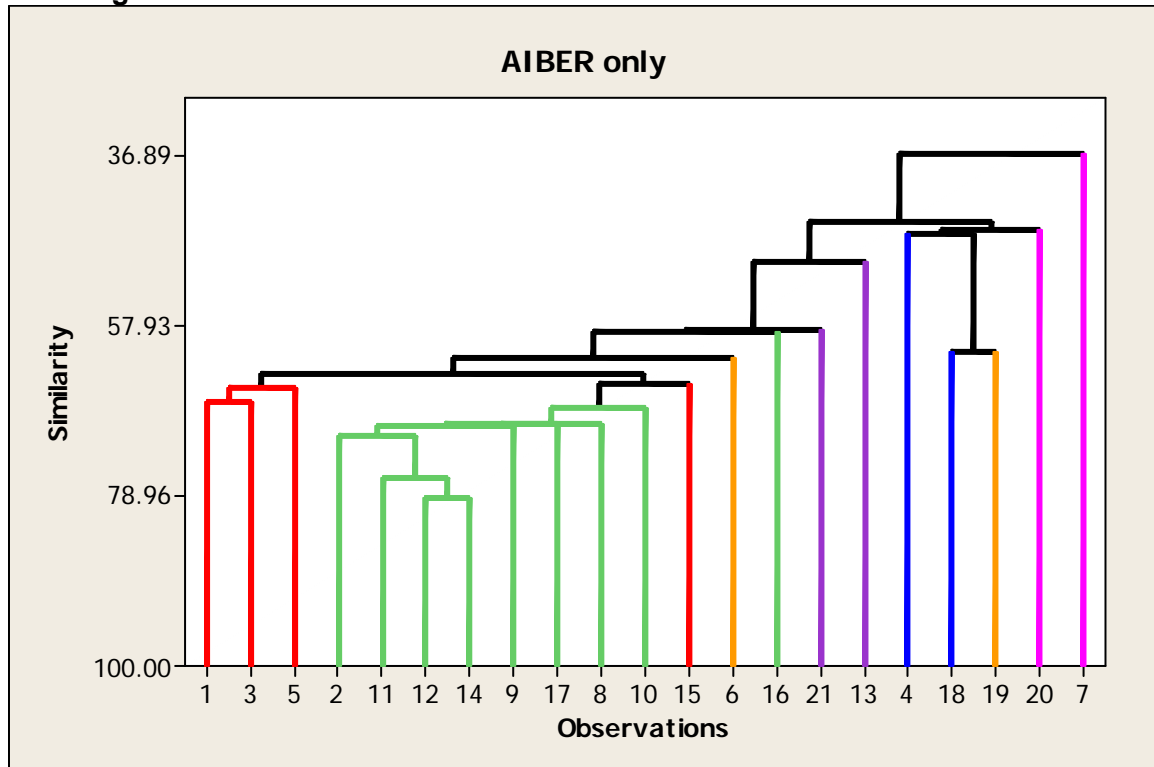
Biodiversity and habitat_n_1	64.360	56.5600	24.620	92.9200
Democracy Index-2014_n_1	74.645	77.7251	79.147	42.4171
Opportunity_n_1	78.681	79.8744	78.885	41.6797
IDH-2014_n_1	80.897	88.6249	74.629	69.5067
Tolerance for immigrants_n_1	75.342	83.5616	84.932	64.3836
EV-Forests_n_1	7.750	45.0500	100.000	33.7600
Water withdrawals_np_1	33.234	25.4877	82.736	54.0016
HPI_n_1	38.814	51.8078	40.367	82.7132
Women treated with respect_n_1	50.000	63.4146	36.585	51.2195
EV-Fisheries_np_1	51.834	62.7743	67.660	70.3668
Health and Wellness_n_1	78.008	92.7093	66.157	73.5374
Religious tolerance_n_1	100.000	66.6667	100.000	66.6667
PC1_1	4.094	4.7767	3.018	0.5190
PC2_1	-1.403	-0.2670	-1.323	-0.5172
PC3_1	0.222	-0.0469	-0.419	2.0777
PC4_1	-0.342	1.1515	-1.827	-0.1809
SPI-2014_n_1	86.071	86.5744	80.715	56.0388

Variable	Grand centroid
EPI-2014_n_1	71.0544
Governança_n_1	42.4840
Biodiversity and habitat_n_1	66.6533
Democracy Index-2014_n_1	59.9752
Opportunity_n_1	55.2807
IDH-2014_n_1	65.6706
Tolerance for immigrants_n_1	61.9700
EV-Forests_n_1	27.2381
Water withdrawals_np_1	62.7811
HPI_n_1	72.3563
Women treated with respect_n_1	26.5720
EV-Fisheries_np_1	56.3900
Health and Wellness_n_1	80.6566
Religious tolerance_n_1	84.1270
PC1_1	1.2384
PC2_1	-1.1879
PC3_1	2.2177
PC4_1	-0.5225
SPI-2014_n_1	64.4072

Distances Between Cluster Centroids

Cluster1	0.000	58.598	109.652	75.717	130.605	73.148	67.202
Cluster2	58.598	0.000	124.517	73.769	112.500	88.809	53.311
Cluster3	109.652	124.517	0.000	105.722	105.486	109.076	121.470
Cluster4	75.717	73.769	105.722	0.000	135.194	109.545	57.749
Cluster5	130.605	112.500	105.486	135.194	0.000	117.280	125.191
Cluster6	73.148	88.809	109.076	109.545	117.280	0.000	99.462
Cluster7	67.202	53.311	121.470	57.749	125.191	99.462	0.000
Cluster8	64.780	78.776	146.020	114.404	152.105	112.879	97.361
Cluster9	99.712	101.123	111.373	83.476	142.590	107.096	95.610
Cluster10	104.060	120.863	81.153	96.323	129.412	98.118	112.369
Cluster11	125.962	132.095	81.844	129.270	124.690	146.748	129.357
Cluster12	75.296	71.682	121.328	100.037	96.069	77.368	82.530
Cluster1	Cluster8	Cluster9	Cluster10	Cluster11	Cluster12		
Cluster1	64.780	99.712	104.060	125.962	75.296		
Cluster2	78.776	101.123	120.863	132.095	71.682		
Cluster3	146.020	111.373	81.153	81.844	121.328		
Cluster4	114.404	83.476	96.323	129.270	100.037		
Cluster5	152.105	142.590	129.412	124.690	96.069		
Cluster6	112.879	107.096	98.118	146.748	77.368		
Cluster7	97.361	95.610	112.369	129.357	82.530		
Cluster8	0.000	121.283	129.696	128.277	103.699		
Cluster9	121.283	0.000	58.871	118.642	117.462		
Cluster10	129.696	58.871	0.000	105.050	112.281		
Cluster11	128.277	118.642	105.050	0.000	143.296		
Cluster12	103.699	117.462	112.281	143.296	0.000		

Dendrogram



A análise por observações corrobora os resultados da análise por variáveis, mostrando que cerca de 52% das observações pertencem aos clusters 1 e 2.

8.2.7 Análise AVECO com 12 clusters (por variáveis e por observações)

Análise AVECO com 12 clusters (por variáveis e por observações) e Dendrogramas

Cluster Analysis of Variables: EPI-2014_n_2, Governança_n, Biodiversity, ...

Correlation Coefficient Distance, Single Linkage
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	18	97.3647	0.052706	5 16	5	2
2	17	97.3147	0.053706	5 6	5	3
3	16	97.1696	0.056608	2 5	2	4
4	15	95.2458	0.095085	2 4	2	5
5	14	94.8526	0.102949	2 8	2	6
6	13	93.6875	0.126250	2 12	2	7
7	12	92.6547	0.146906	14 19	14	2
8	11	91.2472	0.175056	2 7	2	8
9	10	84.2351	0.315297	2 14	2	10
10	9	81.2592	0.374815	2 11	2	11
11	8	79.2532	0.414935	17 18	17	2
12	7	78.6780	0.426440	1 10	1	2
13	6	73.8758	0.522483	1 3	1	3
14	5	73.5142	0.529715	1 2	1	14
15	4	71.7020	0.565960	1 15	1	15
16	3	70.1845	0.596310	1 9	1	16
17	2	67.4238	0.651524	1 17	1	18

18 1 61.6878 0.766243 1 13 1 19

Final Partition

Cluster 1
EPI-2014_n_2

Cluster 2
Governança_n_2 Democracy Index-2014_n_2 SPI-2014_n_2 Opportunity_n_2
Tolerance for immigrants_n_2 Women treated with respect_n_2 PC1_2

Cluster 3
Biodiversity and habitat_n_2

Cluster 4
IDH-2014_n_2

Cluster 5
EV-Forests_n_2

Cluster 6
Water withdrawals_np_2

Cluster 7
HPI_n_2

Cluster 8
EV-Fisheries_np_2

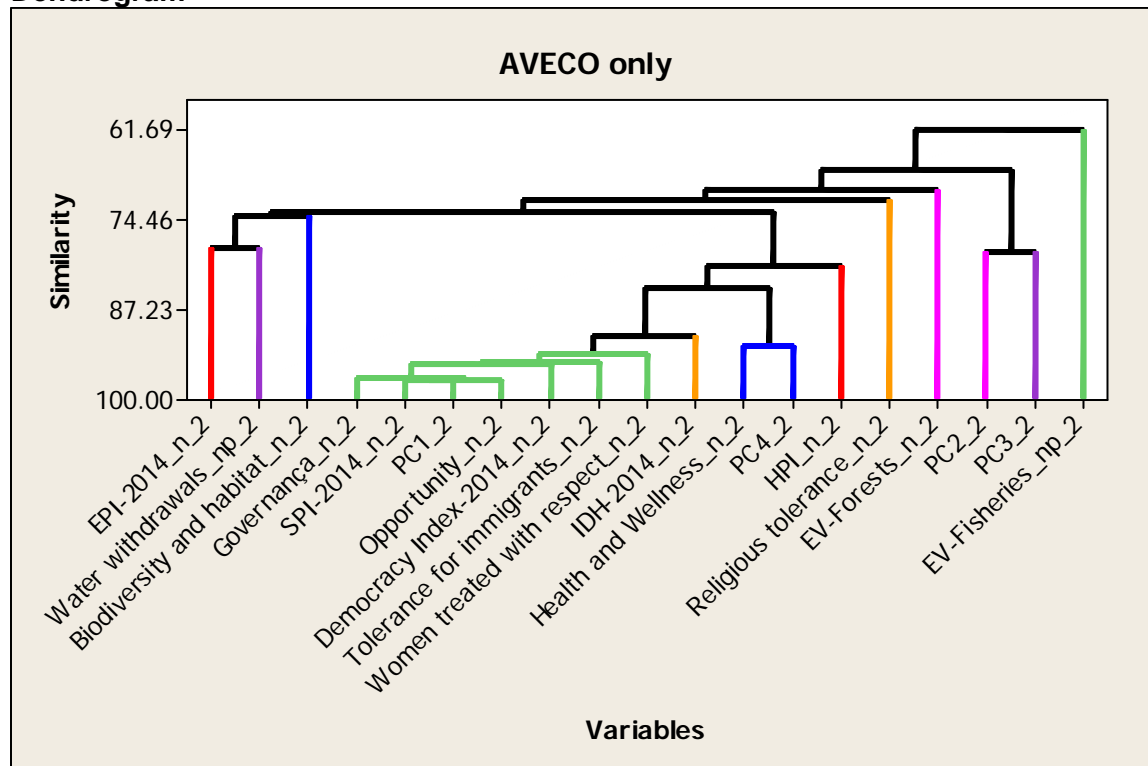
Cluster 9
Health and Wellness_n_2 PC4_2

Cluster 10
Religious tolerance_n_2

Cluster 11
PC2_2

Cluster 12
PC3_2

Dendrogram



No gráfico e tabela acima, identificamos o Cluster 2 como o mais representativo nesta análise, tanto pelo alto grau de similaridade entre as variáveis, quanto pela representatividade, concentrando cerca de 55% das observações. O cluster 2 é composto pelas variáveis:

Governança, Democracy Index-2014, SPI-2014, Opportunity, Tolerance for immigrants, Women treated with respect e PC1.

Cluster Analysis of Observations: EPI-2014_n_2, Governança_n, Biodiversity, ...

Euclidean Distance, Single Linkage
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	26	84.4131	28.1983	5 23	5	2
2	25	83.3805	30.0665	10 19	10	2
3	24	81.1390	34.1216	3 27	3	2
4	23	79.2092	37.6128	2 21	2	2
5	22	78.7265	38.4861	2 25	2	3
6	21	78.2612	39.3279	10 26	10	3
7	20	76.7414	42.0773	9 10	9	4
8	19	76.2541	42.9590	2 8	2	4
9	18	76.0414	43.3436	2 24	2	5
10	17	75.2713	44.7369	2 6	2	6
11	16	73.5693	47.8161	2 5	2	8
12	15	73.0987	48.6674	4 12	4	2
13	14	73.0212	48.8076	2 9	2	12
14	13	72.9790	48.8839	2 4	2	14
15	12	72.2946	50.1222	2 22	2	15
16	11	71.5692	51.4344	15 16	15	2
17	10	70.4866	53.3930	2 3	2	17
18	9	67.2331	59.2789	2 18	2	18
19	8	66.9545	59.7829	11 15	11	3
20	7	66.0333	61.4496	11 17	11	4
21	6	65.1424	63.0613	2 7	2	19
22	5	65.0717	63.1891	2 11	2	23
23	4	63.7257	65.6242	1 20	1	2
24	3	63.1701	66.6293	1 13	1	3
25	2	60.1227	72.1424	1 2	1	26
26	1	55.2126	81.0252	1 14	1	27

Final Partition

Number of clusters: 12

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	1	0.0	0.0000	0.0000
Cluster2	15	35580.4	47.3089	74.5527
Cluster3	2	582.1	17.0608	17.0608
Cluster4	1	0.0	0.0000	0.0000
Cluster5	1	0.0	0.0000	0.0000
Cluster6	1	0.0	0.0000	0.0000
Cluster7	1	0.0	0.0000	0.0000
Cluster8	1	0.0	0.0000	0.0000
Cluster9	1	0.0	0.0000	0.0000
Cluster10	1	0.0	0.0000	0.0000
Cluster11	1	0.0	0.0000	0.0000
Cluster12	1	0.0	0.0000	0.0000

Cluster Centroids

Variable	Cluster1	Cluster2	Cluster3	Cluster4
EPI-2014_n_2	93.542	93.4179	84.6118	96.099
Governança_n_2	91.784	88.2965	83.8541	77.714
Biodiversity and habitat_n_2	83.080	79.4987	60.4550	100.000

Democracy Index-2014_n_2	89.100	85.7109	77.3697	74.052
SPI-2014_n_2	96.154	94.1313	90.0431	87.491
Opportunity_n_2	96.122	82.9105	86.5463	75.369
IDH-2014_n_2	98.503	92.7521	93.0718	86.058
Tolerance for immigrants_n_2	91.781	68.4932	72.6027	20.548
EV-Forests_n_2	100.000	30.7040	18.9950	9.020
Water withdrawals_np_2	29.770	79.0249	39.4717	44.994
HPI_n_2	46.782	52.0484	35.2859	29.809
Women treated with respect_n_2	80.488	75.8537	71.9512	53.659
EV-Fisheries_np_2	68.973	75.4573	97.3250	55.694
Health and Wellness_n_2	95.590	87.8208	77.4077	79.358
Religious tolerance_n_2	66.667	62.2222	66.6667	100.000
PC1_2	6.396	5.5815	5.1284	4.043
PC2_2	-0.557	-0.8508	-0.2685	-0.102
PC3_2	-1.423	-1.0524	-1.3354	0.032
PC4_2	0.928	0.7866	0.3353	-0.326

Variable	Cluster5	Cluster6	Cluster7	Cluster8
EPI-2014_n_2	90.9108	92.385	76.596	88.4285
Governança_n_2	55.2976	87.505	65.156	60.3270
Biodiversity and habitat_n_2	66.4900	18.740	39.340	79.7700
Democracy Index-2014_n_2	70.6161	85.664	72.749	75.3555
SPI-2014_n_2	73.3825	92.469	69.734	79.6729
Opportunity_n_2	53.5950	91.554	51.115	66.3579
IDH-2014_n_2	86.7886	95.243	91.672	88.1042
Tolerance for immigrants_n_2	27.3973	87.671	16.438	68.4932
EV-Forests_n_2	22.8300	100.000	39.720	55.4100
Water withdrawals_np_2	34.5486	41.659	3.492	32.9946
HPI_n_2	43.2719	47.801	78.689	57.3312
Women treated with respect_n_2	57.3171	86.585	53.659	57.3171
EV-Fisheries_np_2	58.5776	73.586	100.000	60.0673
Health and Wellness_n_2	86.1686	93.339	83.228	99.7300
Religious tolerance_n_2	33.3333	66.667	0.000	33.3333
PC1_2	2.8063	6.016	2.983	3.6299
PC2_2	1.5110	-0.080	3.552	0.0864
PC3_2	0.1235	-1.758	0.393	0.2691
PC4_2	0.8865	0.767	1.629	1.3935

Variable	Cluster9	Cluster10	Cluster11	Cluster12
EPI-2014_n_2	81.168	62.542	90.724	94.998
Governança_n_2	84.285	68.293	66.588	98.717
Biodiversity and habitat_n_2	73.530	50.400	95.170	76.050
Democracy Index-2014_n_2	78.081	77.844	70.972	92.062
SPI-2014_n_2	92.757	80.122	74.245	100.000
Opportunity_n_2	85.338	66.358	55.793	100.000
IDH-2014_n_2	91.054	92.354	78.995	94.900
Tolerance for immigrants_n_2	52.055	49.315	19.178	91.781
EV-Forests_n_2	55.410	33.760	3.300	100.000
Water withdrawals_np_2	38.914	29.124	92.968	72.964
HPI_n_2	60.121	51.128	29.626	69.890
Women treated with respect_n_2	54.878	46.341	47.561	79.268
EV-Fisheries_np_2	59.411	64.376	100.000	91.318
Health and Wellness_n_2	100.000	84.098	30.123	83.648
Religious tolerance_n_2	66.667	100.000	100.000	100.000
PC1_2	4.793	3.631	2.368	6.579
PC2_2	0.115	0.470	0.331	-1.223
PC3_2	-0.203	-0.229	-1.207	-1.304
PC4_2	1.047	0.134	-2.270	0.129

Variable	Grand centroid
EPI-2014_n_2	90.2921
Governança_n_2	83.2528
Biodiversity and habitat_n_2	73.9244
Democracy Index-2014_n_2	82.4776
SPI-2014_n_2	90.2994
Opportunity_n_2	79.9389
IDH-2014_n_2	91.8924

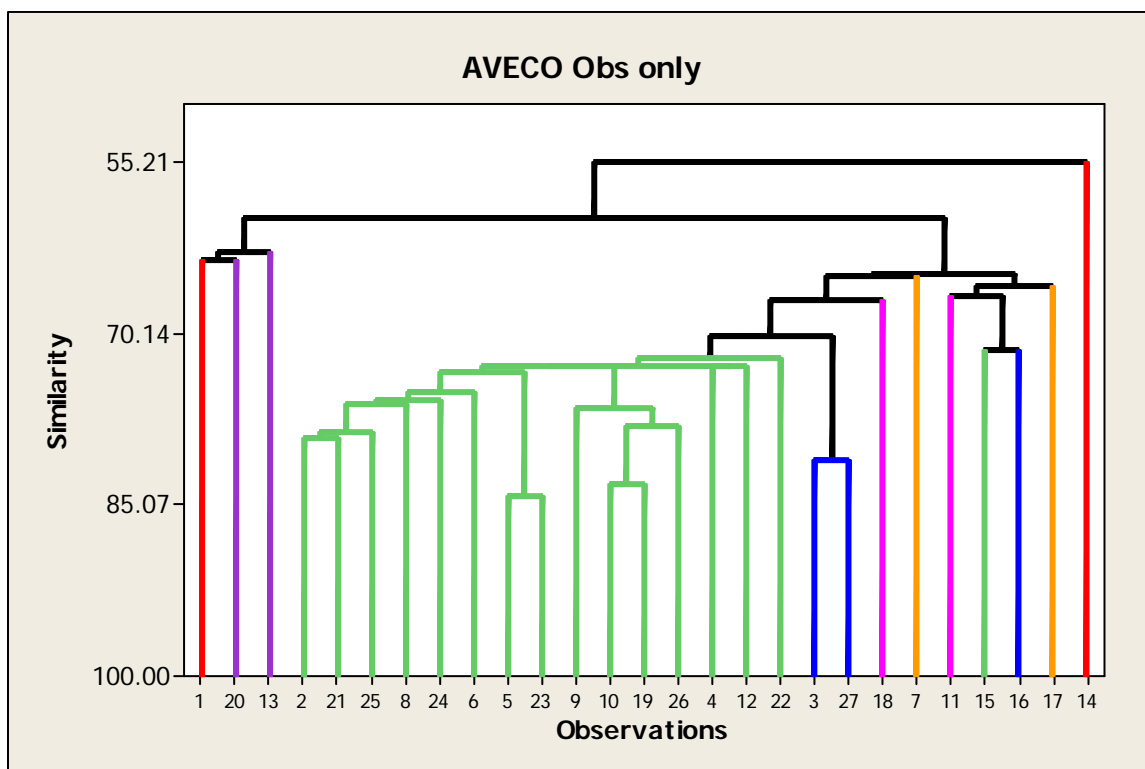
Tolerance for immigrants_n_2	62.8615
EV-Forests_n_2	37.7037
Water withdrawals_np_2	62.4350
HPI_n_2	50.5832
Women treated with respect_n_2	70.3252
EV-Fisheries_np_2	76.2412
Health and Wellness_n_2	85.4597
Religious tolerance_n_2	64.1975
PC1_2	5.0824
PC2_2	-0.3406
PC3_2	-0.8801
PC4_2	0.6217

Distances Between Cluster Centroids

	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6	Cluster7
Cluster1	0.000	90.503	96.519	132.398	128.138	66.629	152.571
Cluster2	90.503	0.000	56.404	87.350	90.024	102.621	135.754
Cluster3	96.519	56.404	0.000	89.336	86.027	99.505	118.984
Cluster4	132.398	87.350	89.336	0.000	85.930	151.841	149.070
Cluster5	128.138	90.024	86.027	85.930	0.000	131.692	81.025
Cluster6	66.629	102.621	99.505	151.841	131.692	0.000	144.844
Cluster7	152.571	135.754	118.984	149.070	81.025	144.844	0.000
Cluster8	83.000	75.762	81.078	105.670	59.783	100.789	97.134
Cluster9	72.142	60.772	69.305	82.405	75.691	89.010	113.095
Cluster10	111.981	89.677	73.190	78.503	82.585	109.798	120.687
Cluster11	173.165	111.003	113.724	86.964	120.309	180.911	167.802
Cluster12	65.624	88.900	107.531	140.978	153.931	81.524	177.449

	Cluster8	Cluster9	Cluster10	Cluster11	Cluster12
Cluster1	83.000	72.142	111.981	173.165	65.624
Cluster2	75.762	60.772	89.677	111.003	88.900
Cluster3	81.078	69.305	73.190	113.724	107.531
Cluster4	105.670	82.405	78.503	86.964	140.978
Cluster5	59.783	75.691	82.585	120.309	153.931
Cluster6	100.789	89.010	109.798	180.911	81.524
Cluster7	97.134	113.095	120.687	167.802	177.449
Cluster8	0.000	51.434	85.676	145.029	117.741
Cluster9	51.434	0.000	61.450	132.836	93.292
Cluster10	85.676	61.450	0.000	117.218	121.081
Cluster11	145.029	132.836	117.218	0.000	159.206
Cluster12	117.741	93.292	121.081	159.206	0.000

Dendrogram



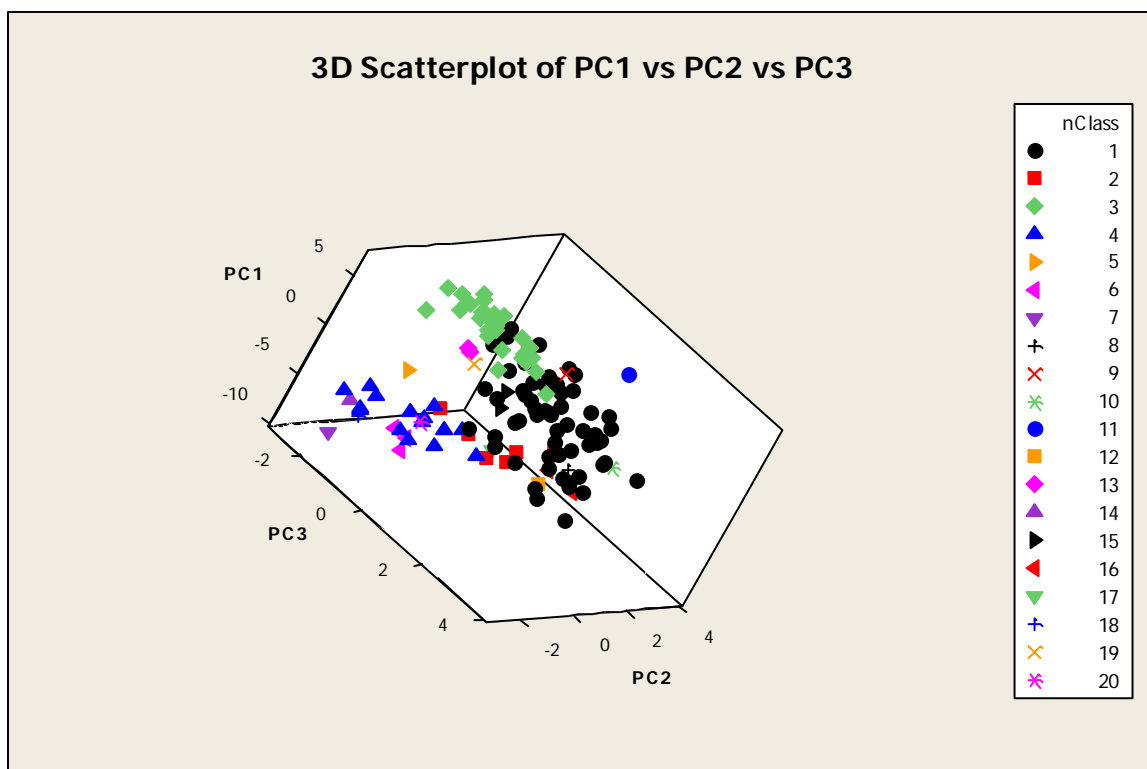
No gráfico e tabela acima, que representa os países do AVECO, identificamos o Cluster 2 como o mais representativo nesta análise, tanto pelo alto grau de similaridade entre as variáveis, quanto pela representatividade, concentrando cerca de 55% das observações. O cluster 2 é composto pelas variáveis: Governança, Democracy Index-2014, SPI-2014, Opportunity, Tolerance for immigrants, Women treated with respect e PC1.

8.3 Gráficos de Dispersão (*Scatterplot 3D*)

Análise em formato de diagramas de dispersão (*scatterplot*), tomando como referência os 4 PCs e amostras de todos os países, somente AIBER e somente AVECO

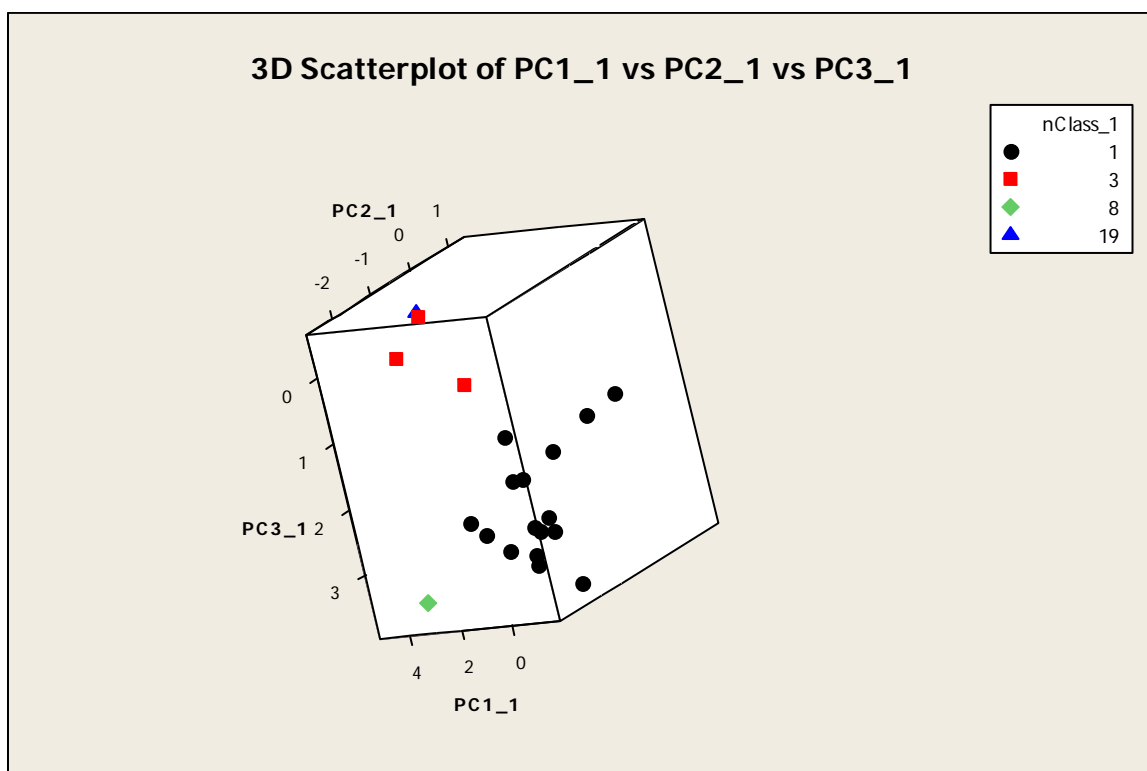
8.3.1 Gráficos de Dispersão (*scatterplot*) – PCs e todos os países e regiões

Análise dos gráficos de dispersão tomando como base os PCs e todos os países e regiões.



8.3.2 Gráficos de Dispersão (*scatterplot*) – PCs e a região AIBER

Análise dos gráficos de dispersão tomando como base os PCs e a região AIBER.

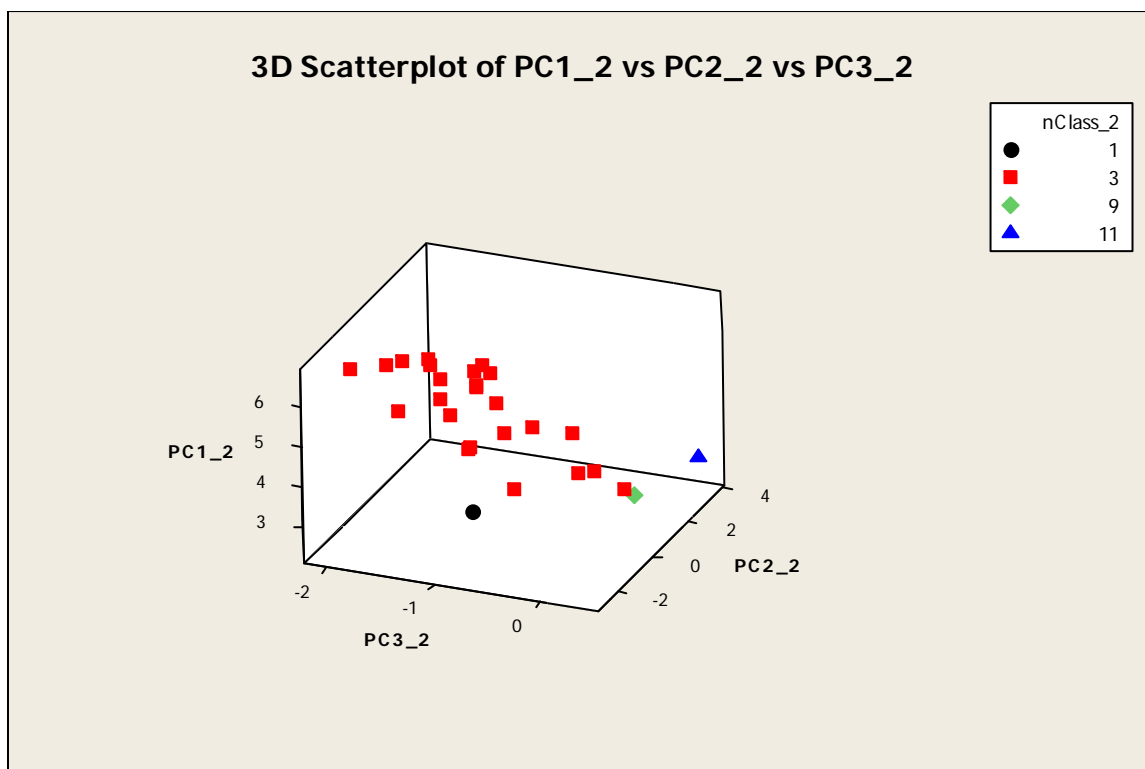


Neste gráfico de dispersão utilizando a coluna Nova Classe (nClass_1) como variável categórica, dos 21 países AIBER analisados destacaram-se Costa Rica (verde), Uruguai

(azul), Portugal, Espanha e Chile (vermelho). Enquanto os demais permaneceram agrupados (cor preta).

8.3.3 Gráficos de Dispersão (*scatterplot*) – PCs e a região AVECO

Análise dos gráficos de dispersão tomando como base os PCs e a região AVECO.



Neste gráfico de dispersão utilizando a coluna Nova Classe (nClass_2) como variável categórica obtivemos, dos 27 países AVECO analisados destacaram-se Letônia (preto), Grécia (verde) e Israel (azul). Enquanto os demais permaneceram agrupados (cor vermelha).

8.4. CONSIDERAÇÕES

Nesta sessão foram realizadas análises de conglomerados, buscando identificar agrupamentos por similaridade. E análise de dendrogramas e clusters dos dados.

Para estabelecer a quantidade de clusters que melhor explicasse os dados deste estudo, testou-se as quantidades de 1, 10, 12, 20 e 30 clusters entre todas as regiões e somente AIBER e somente AVECO e com todas as variáveis e somente com os PCs. A configuração de 12 clusters refeltiu melhor a situação das três, entre todas estas combinações, por isso foi a adotada.

Durante a execução das análises de cada região e dos PCs, notou-se que o modelo de clusterização escolhido não produziu o efeito esperado, tanto que na região AIBER mais de 50% das variáveis se concentraram no cluster 1, restando apenas 1 variável para os demais clusters. E na quase totalidade dos casos, percebemos uma grande concentração do número de observações numa pequena quantidade de clusters. Não foi possível identificar tendências através dos centróides, uma vez que não apresentaram proporcionalidade nos resultados das colunas de cluster.

9. ANÁLISE DISCRIMINANTE

Esta sessão tratará de algumas ferramentas e técnicas como análise de regressão, a fim de identificar as variáveis que melhor explicam os dados deste estudo, agrupamentos, de maneira que se tenha uma quantidade reduzida de variáveis com o melhor grau de explicação dos resultados. Para em seguida realizar análise discriminante para identificar o grau de acerto na distribuição das observações em seus respectivos grupos.

9.1. Análises de Regressão, de Agrupamentos (cluster) e Discriminante

As variáveis para este trabalho, conforme detalhadas na sessão anterior, serão: EPI-2014, Governança, Biodiversity and habitat, Democracy Index-2014, SPI-2014, Opportunity, IDH-2014, Tolerance for immigrants, EV-Forests, Water withdrawals, HPI, Women treated with respect, EV-Fisheries, Health and Wellness, Religious tolerance.

Utilizamos da função “General Regression” tendo SPI como “Response”, para analisar a equação abaixo:

General Regression Analysis: SPI-2014_n versus Opportunity, Water withdrawals, ...

Regression Equation

$$\text{SPI-2014_n} = -3.18942 + 0.507796 \text{ Opportunity_n} - 0.00943243 \text{ Water withdrawals_np} + 0.0491839 \text{ Biodiversity and habitat_n} + 0.0507202 \text{ Women treated with respect_n} - 0.0470843 \text{ Tolerance for immigrants_n} - 0.00713546 \text{ Religious tolerance_n} + 0.238379 \text{ Governança_n} + 0.339865 \text{ EPI-2014_n} + 0.0235241 \text{ EV-Forests_n} + 0.0104549 \text{ EV-Fisheries_np} + 0.118185 \text{ HPI_n} - 0.101345 \text{ Democracy Index-2014_n}$$

Coefficients

Term	Coef	SE Coef	T	P
Constant	-3.18942	4.03399	-0.79064	0.431
Opportunity_n	0.50780	0.08149	6.23110	0.000
Water withdrawals_np	-0.00943	0.01780	-0.52984	0.597

Biodiversity and habitat_n	0.04918	0.02088	2.35608	0.020
Women treated with respect_n	0.05072	0.02702	1.87730	0.063
Tolerance for immigrants_n	-0.04708	0.03036	-1.55091	0.124
Religious tolerance_n	-0.00714	0.01767	-0.40391	0.687
Governança_n	0.23838	0.06194	3.84867	0.000
EPI-2014_n	0.33987	0.03549	9.57751	0.000
EV-Forests_n	0.02352	0.01962	1.19916	0.233
EV-Fisheries_np	0.01045	0.02846	0.36735	0.714
HPI_n	0.11818	0.02606	4.53513	0.000
Democracy Index-2014_n	-0.10134	0.04998	-2.02765	0.045

Summary of Model

S = 5.48267 R-Sq = 95.81% R-Sq(adj) = 95.39%
PRESS = 4471.62 R-Sq(pred) = 94.76%

Analysis of Variance

Source	DF	Seq SS	Adj SS	Adj MS	F
Regression	12	81755.9	81755.9	6812.99	226.649
Opportunity_n	1	72531.5	1167.1	1167.12	38.827
Water withdrawals_np	1	983.2	8.4	8.44	0.281
Biodiversity and habitat_n	1	527.0	166.9	166.86	5.551
Women treated with respect_n	1	124.3	105.9	105.94	3.524
Tolerance for immigrants_n	1	2065.2	72.3	72.30	2.405
Religious tolerance_n	1	351.9	4.9	4.90	0.163
Governança_n	1	237.6	445.3	445.25	14.812
EPI-2014_n	1	4214.3	2757.3	2757.33	91.729
EV-Forests_n	1	34.2	43.2	43.22	1.438
EV-Fisheries_np	1	0.1	4.1	4.06	0.135
HPI_n	1	562.9	618.2	618.25	20.567
Democracy Index-2014_n	1	123.6	123.6	123.59	4.111
Error	119	3577.1	3577.1	30.06	
Total	131	85333.0			

Source	P
Regression	0.000000
Opportunity_n	0.000000
Water withdrawals_np	0.597211
Biodiversity and habitat_n	0.020104
Women treated with respect_n	0.062925
Tolerance for immigrants_n	0.123578
Religious tolerance_n	0.687006
Governança_n	0.000193
EPI-2014_n	0.000000
EV-Forests_n	0.232850
EV-Fisheries_np	0.714007
HPI_n	0.000014
Democracy Index-2014_n	0.044829
Error	
Total	

Fits and Diagnostics for Unusual Observations

Obs	SPI-2014_n	Fit	SE Fit	Residual	St Resid	
14	58.2135	44.0759	1.95875	14.1376	2.76080	R
37	49.1912	38.0216	1.57593	11.1696	2.12703	R
64	80.1222	69.1822	1.35236	10.9400	2.05900	R
65	68.4040	52.5324	2.02685	15.8717	3.11559	R
113	29.2416	40.2637	1.37335	-11.0222	-2.07657	R
132	31.0568	45.5732	1.37424	-14.5164	-2.73500	R

R denotes an observation with a large standardized residual.

Alterando variável de Response de SPI para IDH:

General Regression Analysis: IDH-2014_n versus Opportunity, Water withdrawals, ...

Regression Equation

IDH-2014_n = 14.6144 + 0.365813 Opportunity_n - 0.0977047 Water withdrawals_np + 0.00925913 Biodiversity and habitat_n + 0.0142142 Women treated with respect_n - 0.106155 Tolerance for immigrants_n - 0.0776389 Religious tolerance_n + 0.255818 Governança_n + 0.539926 EPI-2014_n + 0.0204428 EV-Forests_n + 0.0533983 EV-Fisheries_np + 0.0927233 HPI_n - 0.200201 Democracy Index-2014_n

Coefficients

Term	Coef	SE Coef	T	P
Constant	14.6144	6.31033	2.31594	0.022
Opportunity_n	0.3658	0.12748	2.86958	0.005
Water withdrawals_np	-0.0977	0.02785	-3.50847	0.001
Biodiversity and habitat_n	0.0093	0.03266	0.28354	0.777
Women treated with respect_n	0.0142	0.04226	0.33633	0.737
Tolerance for immigrants_n	-0.1062	0.04749	-2.23529	0.027
Religious tolerance_n	-0.0776	0.02763	-2.80945	0.006
Governança_n	0.2558	0.09689	2.64032	0.009
EPI-2014_n	0.5399	0.05551	9.72664	0.000
EV-Forests_n	0.0204	0.03069	0.66617	0.507
EV-Fisheries_np	0.0534	0.04452	1.19944	0.233
HPI_n	0.0927	0.04077	2.27457	0.025
Democracy Index-2014_n	-0.2002	0.07819	-2.56059	0.012

Summary of Model

S = 8.57648 R-Sq = 90.44% R-Sq(adj) = 89.48%
 PRESS = 11007.1 R-Sq(pred) = 87.98%

Analysis of Variance

Source	DF	Seq SS	Adj SS	Adj MS	F
Regression	12	82815.4	82815.4	6901.28	93.8235
Opportunity_n	1	56512.3	605.7	605.70	8.2345
Water withdrawals_np	1	7027.5	905.4	905.43	12.3094
Biodiversity and habitat_n	1	215.3	5.9	5.91	0.0804
Women treated with respect_n	1	41.8	8.3	8.32	0.1131
Tolerance for immigrants_n	1	5698.9	367.5	367.52	4.9965
Religious tolerance_n	1	2082.4	580.6	580.58	7.8930
Governança_n	1	183.6	512.8	512.78	6.9713
EPI-2014_n	1	10122.0	6959.0	6958.95	94.6075
EV-Forests_n	1	52.0	32.6	32.64	0.4438
EV-Fisheries_np	1	106.6	105.8	105.82	1.4387
HPI_n	1	290.9	380.6	380.56	5.1737
Democracy Index-2014_n	1	482.3	482.3	482.28	6.5566
Error	119	8753.2	8753.2	73.56	
Total	131	91568.6			

Source	P
Regression	0.000000
Opportunity_n	0.004865
Water withdrawals_np	0.000637
Biodiversity and habitat_n	0.777253
Women treated with respect_n	0.737218
Tolerance for immigrants_n	0.027264
Religious tolerance_n	0.005803
Governança_n	0.009393
EPI-2014_n	0.000000
EV-Forests_n	0.506592
EV-Fisheries_np	0.232741

HPI_n	0.024723
Democracy Index-2014_n	0.011700
Error	
Total	

Fits and Diagnostics for Unusual Observations

Obs	IDH-2014_n	Fit	SE Fit	Residual	St Resid	
64	92.3537	68.6928	2.11548	23.6609	2.84678	R
78	26.4618	10.3259	3.36911	16.1359	2.04588	R
84	46.9591	65.8935	2.74933	-18.9344	-2.33071	R
105	19.7499	36.9712	2.70060	-17.2213	-2.11559	R
113	30.6198	49.6268	2.14832	-19.0070	-2.28916	R
116	46.3718	63.5185	3.11384	-17.1467	-2.14568	R

R denotes an observation with a large standardized residual.

Com a alteração da variável de “response” de SPI para IDH, houve uma redução no grau de explicação (r quadrado) de 95 para 90%, portanto utilizaremos o IDH para as análises a seguir.

Iniciar processo de retirar variáveis a ver se equilibra a proporcionalidade, mas mantém r^2 próximo do inicial (+ou- 90%).

Excluindo as seguintes variáveis que apresentaram p-value > 0.05

Biodiversity and habitat_n	0.0093	0.03266	0.28354	0.777
Women treated with respect_n	0.0142	0.04226	0.33633	0.737
EV-Forests_n	0.0204	0.03069	0.66617	0.507
EV-Fisheries_np	0.0534	0.04452	1.19944	0.233
Health and Wellness_n	0.0854	0.05288	1.61442	0.109

Executar a regressão novamente considerando as seguintes variáveis:

Term	Coef	SE Coef	T	P
Constant	14.6144	6.31033	2.31594	0.022
Opportunity_n	0.3658	0.12748	2.86958	0.005
Water withdrawals_np	-0.0977	0.02785	-3.50847	0.001
Tolerance for immigrants_n	-0.1062	0.04749	-2.23529	0.027
Religious tolerance_n	-0.0776	0.02763	-2.80945	0.006
Governança_n	0.2558	0.09689	2.64032	0.009
EPI-2014_n	0.5399	0.05551	9.72664	0.000
HPI_n	0.0927	0.04077	2.27457	0.025
Democracy Index-2014_n	-0.2002	0.07819	-2.56059	0.012

Resultados da regressão:

General Regression Analysis: IDH-2014_n versus Opportunity, Water withdrawals, ...

Regression Equation

$$\text{IDH-2014}_n = 21.1823 + 0.368822 \text{ Opportunity}_n - 0.0991802 \text{ Water withdrawals}_np - 0.105761 \text{ Tolerance for immigrants}_n - 0.0820697 \text{ Religious tolerance}_n + 0.290008 \text{ Governança}_n + 0.539847 \text{ EPI-2014}_n + 0.0816496 \text{ HPI}_n - 0.227971 \text{ Democracy Index-2014}_n$$

Coefficients

Term	Coef	SE Coef	T	P
Constant	21.1823	3.89519	5.4381	0.000
Opportunity_n	0.3688	0.12223	3.0174	0.003
Water withdrawals_np	-0.0992	0.02685	-3.6943	0.000
Tolerance for immigrants_n	-0.1058	0.04561	-2.3186	0.022
Religious tolerance_n	-0.0821	0.02661	-3.0844	0.003
Governança_n	0.2900	0.08230	3.5239	0.001
EPI-2014_n	0.5398	0.05042	10.7065	0.000
HPI_n	0.0816	0.03966	2.0585	0.042
Democracy Index-2014_n	-0.2280	0.07112	-3.2056	0.002

Summary of Model

S = 8.51629 R-Sq = 90.26% R-Sq(adj) = 89.62%
 PRESS = 10414.3 R-Sq(pred) = 88.63%

Analysis of Variance

Source	DF	Seq SS	Adj SS	Adj MS	F
Regression	8	82647.7	82647.7	10331.0	142.443
Opportunity_n	1	56512.3	660.3	660.3	9.105
Water withdrawals_np	1	7027.5	989.8	989.8	13.648
Tolerance for immigrants_n	1	5646.3	389.9	389.9	5.376
Religious tolerance_n	1	2109.0	690.0	690.0	9.513
Governança_n	1	254.2	900.6	900.6	12.418
EPI-2014_n	1	10205.9	8313.8	8313.8	114.630
HPI_n	1	147.3	307.3	307.3	4.238
Democracy Index-2014_n	1	745.3	745.3	745.3	10.276
Error	123	8920.8	8920.8	72.5	
Total	131	91568.6			

Source	P
Regression	0.0000000
Opportunity_n	0.0030994
Water withdrawals_np	0.0003303
Tolerance for immigrants_n	0.0220676
Religious tolerance_n	0.0025192
Governança_n	0.0005979
EPI-2014_n	0.0000000
HPI_n	0.0416509
Democracy Index-2014_n	0.0017172
Error	
Total	

Fits and Diagnostics for Unusual Observations

Obs	IDH-2014_n	Fit	SE Fit	Residual	St Resid	
64	92.3537	69.3427	1.99433	23.0110	2.77928	R
65	78.5761	61.9678	2.66048	16.6084	2.05294	R
78	26.4618	10.3997	3.01358	16.0622	2.01653	R
84	46.9591	65.2777	2.42454	-18.3186	-2.24387	R
105	19.7499	36.8387	2.62333	-17.0888	-2.10916	R
113	30.6198	50.2481	1.67926	-19.6282	-2.35094	R
116	46.3718	63.2430	2.52604	-16.8712	-2.07441	R

R denotes an observation with a large standardized residual.

Aparentemente obtivemos um resultado satisfatório com esta composição de variáveis. O coeficiente de determinação (r-quadrado) se manteve na ordem de 90% e o p-value < 0.05.

Mesmo com várias combinações, não foi possível melhorar a proporcionalidade (peso) das variáveis na equação final.

Utilizaremos as mesmas variáveis da regressão para criar agrupamentos de similaridade.

Executar a função Cluster Obs, com as variáveis:

'Opportunity_n' 'Water withdrawals_np' 'Tolerance for immigrants_n' 'Religious tolerance_n'

'Governança_n' 'EPI-2014_n' 'HPI_n' 'Democracy Index-2014_n' 'IDH-2014_n'

Utilizar os parâmetros:

Linkage Method: single

Distance measure: Manhattan

Nr of cluster: 4

Cluster Analysis of Observations: Opportunity, Water withdrawals, Tolerance for Im, ...

Manhattan Distance, Single Linkage
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	131	95.3222	31.821	35 97	35	2
2	130	95.2452	32.345	56 127	56	2
3	129	95.0434	33.717	6 56	6	3
4	128	94.5520	37.060	68 107	68	2
5	127	94.4723	37.602	93 115	93	2
6	126	94.3038	38.748	33 40	33	2
7	125	94.1522	39.780	71 85	71	2
8	124	93.9747	40.987	43 88	43	2
9	123	93.7851	42.277	46 49	46	2
10	122	93.7736	42.356	43 126	43	3
11	121	93.7357	42.613	72 99	72	2
12	120	93.7288	42.660	32 72	32	3
13	119	93.6954	42.887	17 101	17	2
14	118	93.5863	43.630	11 60	11	2
15	117	93.5066	44.171	36 48	36	2
16	116	93.3775	45.050	33 93	33	4
17	115	93.3023	45.561	22 89	22	2
18	114	93.0428	47.327	19 119	19	2
19	113	92.6992	49.664	41 43	41	4
20	112	92.5858	50.435	46 90	46	3
21	111	92.5359	50.775	41 114	41	5
22	110	92.2519	52.707	32 108	32	4
23	109	92.2292	52.861	32 50	32	5
24	108	92.0678	53.959	18 71	18	3
25	107	92.0107	54.348	17 83	17	3
26	106	92.0011	54.413	1 36	1	3
27	105	91.9677	54.640	12 18	12	4
28	104	91.5875	57.227	19 21	19	3
29	103	91.5277	57.633	5 122	5	2
30	102	91.3053	59.146	5 42	5	3
31	101	91.2231	59.705	6 11	6	5
32	100	91.0282	61.031	116 129	116	2

33	99	91.0262	61.045	12	105	12	5
34	98	90.9845	61.328	7	33	7	5
35	97	90.9715	61.417	12	132	12	6
36	96	90.9698	61.429	1	35	1	5
37	95	90.9201	61.766	1	46	1	8
38	94	90.9053	61.867	1	38	1	9
39	93	90.8843	62.010	5	121	5	4
40	92	90.8471	62.263	1	86	1	10
41	91	90.7767	62.742	1	98	1	11
42	90	90.4941	64.664	6	25	6	6
43	89	90.4914	64.683	32	68	32	7
44	88	90.3854	65.404	16	106	16	2
45	87	90.2789	66.129	19	78	19	4
46	86	90.2225	66.512	2	26	2	2
47	85	90.1438	67.047	5	8	5	5
48	84	89.9559	68.326	30	32	30	8
49	83	89.8824	68.826	5	81	5	6
50	82	89.8404	69.111	5	111	5	7
51	81	89.7982	69.398	1	13	1	12
52	80	89.6163	70.636	6	41	6	11
53	79	89.5484	71.097	4	16	4	3
54	78	89.4107	72.034	19	28	19	5
55	77	89.3852	72.207	6	110	6	12
56	76	89.2439	73.169	6	7	6	17
57	75	89.2382	73.208	84	116	84	3
58	74	89.2323	73.248	5	80	5	8
59	73	89.2307	73.259	23	47	23	2
60	72	89.1611	73.732	2	5	2	10
61	71	89.1454	73.839	12	91	12	7
62	70	88.7711	76.385	12	19	12	12
63	69	88.7404	76.594	6	30	6	25
64	68	88.6017	77.537	12	75	12	13
65	67	88.5329	78.006	1	4	1	15
66	66	88.4367	78.660	73	109	73	2
67	65	88.4294	78.710	6	100	6	26
68	64	88.3745	79.083	6	58	6	27
69	63	88.2875	79.675	2	53	2	11
70	62	88.2731	79.773	2	14	2	12
71	61	88.1950	80.304	12	103	12	14
72	60	88.1569	80.563	12	44	12	15
73	59	88.1411	80.671	63	92	63	2
74	58	88.1360	80.706	22	51	22	3
75	57	88.1044	80.920	2	118	2	13
76	56	87.9110	82.236	12	77	12	16
77	55	87.8902	82.378	117	123	117	2
78	54	87.8509	82.645	1	17	1	18
79	53	87.8318	82.775	1	6	1	45
80	52	87.8009	82.985	2	52	2	14
81	51	87.7930	83.039	1	45	1	46
82	50	87.7001	83.671	2	12	2	30
83	49	87.6920	83.726	1	2	1	76
84	48	87.5094	84.968	1	22	1	79
85	47	87.4574	85.322	1	61	1	80
86	46	87.3431	86.099	55	94	55	2
87	45	87.1832	87.187	3	117	3	3
88	44	87.1611	87.337	1	69	1	81
89	43	87.0065	88.389	1	39	1	82
90	42	86.9615	88.695	1	66	1	83
91	41	86.8457	89.483	1	95	1	84
92	40	86.6991	90.480	1	124	1	85
93	39	86.4427	92.224	3	23	3	5
94	38	86.3904	92.580	54	104	54	2
95	37	86.3558	92.816	79	128	79	2
96	36	86.3346	92.960	1	84	1	88
97	35	86.1765	94.035	3	63	3	7
98	34	86.0566	94.851	59	120	59	2
99	33	85.9077	95.864	1	24	1	89
100	32	85.8825	96.035	1	96	1	90

Final Partition

Number of clusters: 4

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	129	802305	76.1081	126.548
Cluster2	1	0	0.0000	0.000
Cluster3	1	0	0.0000	0.000
Cluster4	1	0	0.0000	0.000

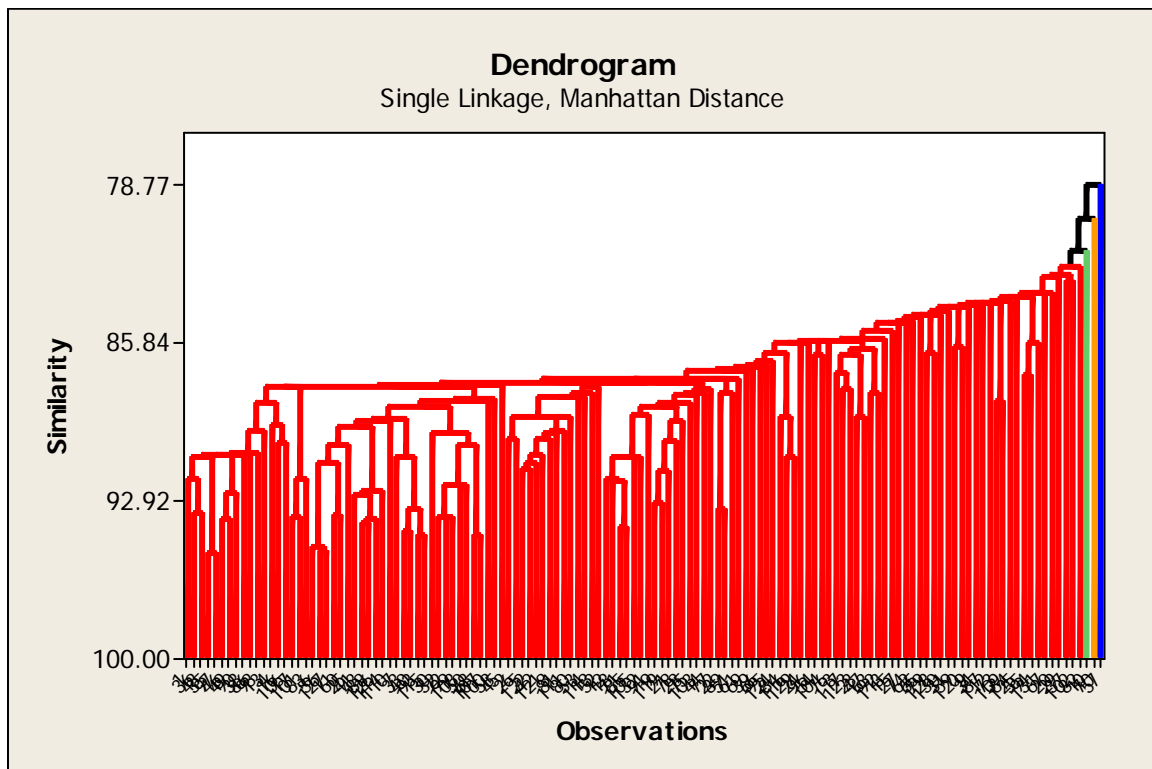
Cluster Centroids

Variable	Cluster1	Cluster2	Cluster3	Cluster4	Grand centroid
Opportunity_n	45.8783	31.115	51.1146	44.364	45.7947
Water withdrawals_np	61.9068	88.172	3.4923	20.547	61.3499
Tolerance for immigrants_n	52.3521	35.616	16.4384	32.877	51.8057
Religious tolerance_n	59.1731	100.000	0.0000	100.000	59.3434
Governança_n	45.0480	21.841	65.1564	44.183	45.0180
EPI-2014_n	60.8271	84.360	76.5957	18.794	60.8064
HPI_n	48.3182	35.767	78.6888	48.454	48.4542
Democracy Index-2014_n	51.9003	26.066	72.7488	61.256	51.9334
IDH-2014_n	59.5345	75.570	91.6719	24.895	59.6370

Distances Between Cluster Centroids

	Cluster1	Cluster2	Cluster3	Cluster4
Cluster1	0.000	70.916	106.178	82.547
Cluster2	70.916	0.000	155.421	116.282
Cluster3	106.178	155.421	0.000	141.042
Cluster4	82.547	116.282	141.042	0.000

Dendrogram



No dendrograma acima, percebemos uma grande concentração de observações agrupadas no cluster 1 (total de 129 obs), portanto o método “single” não conseguiu distribuir os dados de maneira satisfatória.

Executar novamente, porém utilizando “Linkage Method” como “Complete”, demais parâmetros permanecem os mesmos. Distance measure: Manhattan and Nr of clusters: 4. Gerar coluna de “Cluster Membership column”.

Resultados:

Cluster Analysis of Observations: Opportunity, Water withdrawals, Tolerance for Im, ...

Manhattan Distance, Complete Linkage
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	131	95.3222	31.821	35 97	35	2
2	130	95.2452	32.345	56 127	56	2
3	129	94.5520	37.060	68 107	68	2
4	128	94.4723	37.602	93 115	93	2
5	127	94.3038	38.748	33 40	33	2
6	126	94.1522	39.780	71 85	71	2
7	125	93.9747	40.987	43 88	43	2
8	124	93.7851	42.277	46 49	46	2
9	123	93.7357	42.613	72 99	72	2
10	122	93.6954	42.887	17 101	17	2
11	121	93.5863	43.630	11 60	11	2
12	120	93.5066	44.171	36 48	36	2
13	119	93.3023	45.561	22 89	22	2
14	118	93.0428	47.327	19 119	19	2
15	117	92.6692	49.868	32 72	32	3
16	116	91.9677	54.640	12 18	12	2
17	115	91.8450	55.475	1 36	1	3
18	114	91.5277	57.633	5 122	5	2
19	113	91.1426	60.253	50 108	50	2
20	112	91.0282	61.031	116 129	116	2
21	111	90.9715	61.417	105 132	105	2
22	110	90.8835	62.015	12 71	12	4
23	109	90.8815	62.029	46 90	46	3
24	108	90.7784	62.730	43 126	43	3
25	107	90.5882	64.024	6 56	6	3
26	106	90.3854	65.404	16 106	16	2
27	105	90.2225	66.512	2 26	2	2
28	104	90.0139	67.931	35 98	35	3
29	103	89.8824	68.826	42 81	42	2
30	102	89.7863	69.479	43 114	43	4
31	101	89.4581	71.712	33 93	33	4
32	100	89.4058	72.068	19 28	19	3
33	99	89.3806	72.239	5 121	5	3
34	98	89.2307	73.259	23 47	23	2
35	97	89.0005	74.824	11 25	11	3
36	96	88.4367	78.660	73 109	73	2
37	95	88.4294	78.710	100 110	100	2
38	94	88.3745	79.083	41 58	41	2
39	93	88.3179	79.468	17 83	17	3
40	92	88.2875	79.675	53 111	53	2
41	91	88.2049	80.237	30 68	30	3
42	90	88.1950	80.304	21 103	21	2
43	89	88.1411	80.671	63 92	63	2

44	88	87.9994	81.635	2	8	2	3
45	87	87.8902	82.378	117	123	117	2
46	86	87.8150	82.889	1	38	1	4
47	85	87.7001	83.671	14	44	14	2
48	84	87.3431	86.099	55	94	55	2
49	83	86.9522	88.758	13	46	13	4
50	82	86.8878	89.197	84	116	84	3
51	81	86.8479	89.468	12	91	12	5
52	80	86.7637	90.041	7	33	7	5
53	79	86.6600	90.746	32	50	32	5
54	78	86.5977	91.170	61	66	61	2
55	77	86.3904	92.580	54	104	54	2
56	76	86.3558	92.816	79	128	79	2
57	75	86.2958	93.224	21	78	21	3
58	74	86.2930	93.242	53	118	53	3
59	73	86.0951	94.589	22	51	22	3
60	72	86.0566	94.851	59	120	59	2
61	71	85.2915	100.056	15	86	15	2
62	70	85.1771	100.834	75	105	75	3
63	69	85.1433	101.064	12	77	12	6
64	68	85.0879	101.440	42	76	42	3
65	67	84.9464	102.403	45	80	45	2
66	66	84.5826	104.877	4	130	4	2
67	65	84.3974	106.138	19	21	19	6
68	64	84.0719	108.352	74	87	74	2
69	63	83.8958	109.550	16	27	16	3
70	62	83.8768	109.679	84	113	84	4
71	61	83.8137	110.108	65	125	65	2
72	60	83.5811	111.690	1	13	1	8
73	59	83.3732	113.105	2	5	2	6
74	58	83.1494	114.627	42	124	42	4
75	57	83.1051	114.928	37	102	37	2
76	56	82.4508	119.379	30	32	30	8
77	55	82.2748	120.576	11	41	11	5
78	54	82.0321	122.227	62	84	62	5
79	53	81.9344	122.892	17	42	17	7
80	52	81.7334	124.259	29	95	29	2
81	51	81.6949	124.521	10	31	10	2
82	50	81.5872	125.254	64	100	64	3
83	49	81.3959	126.556	2	61	2	8
84	48	80.8416	130.326	67	74	67	3
85	47	80.7901	130.676	55	131	55	3
86	46	80.7764	130.769	7	43	7	9
87	45	80.6964	131.314	3	112	3	2
88	44	80.6192	131.839	23	24	23	3
89	43	80.5349	132.412	73	82	73	3
90	42	80.4268	133.147	16	96	16	4
91	41	80.3770	133.486	2	69	2	9
92	40	79.8127	137.325	34	70	34	2
93	39	79.5640	139.017	14	117	14	4
94	38	79.4405	139.857	4	35	4	5
95	37	78.6634	145.143	6	7	6	12
96	36	78.4887	146.332	45	57	45	3
97	35	78.3819	147.058	3	63	3	4
98	34	78.0425	149.367	64	79	64	5
99	33	77.8832	150.451	52	53	52	4
100	32	77.6015	152.367	15	73	15	5

Final Partition

Number of clusters: 4

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	28	74592	49.0583	78.925
Cluster2	34	96312	50.7613	87.392
Cluster3	36	131979	57.7593	109.353
Cluster4	34	77881	46.7241	69.964

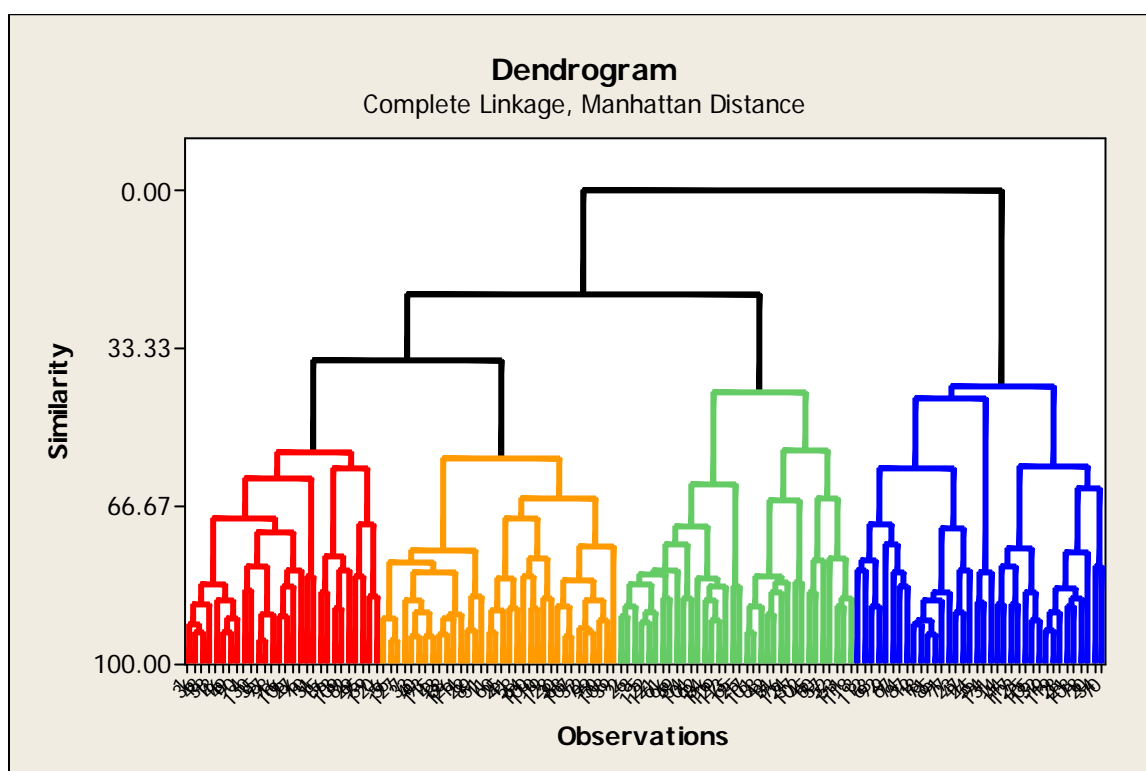
Cluster Centroids

Variable	Cluster1	Cluster2	Cluster3	Cluster4	Grand centroid
Opportunity_n	50.4906	34.9307	21.7168	78.2856	45.7947
Water withdrawals_np	60.2911	36.7576	81.5041	65.4743	61.3499
Tolerance for immigrants_n	56.1155	34.3674	53.9193	63.4569	51.8057
Religious tolerance_n	88.0952	31.3725	53.7037	69.6078	59.3434
Governança_n	38.8391	36.7271	23.6891	80.9810	45.0180
EPI-2014_n	68.1924	65.6242	23.9045	88.9785	60.8064
HPI_n	61.4019	52.7796	34.4957	48.2457	48.4542
Democracy Index-2014_n	57.2571	38.9044	33.0338	80.5896	51.9334
IDH-2014_n	61.4657	66.1259	24.3879	88.9648	59.6370

Distances Between Cluster Centroids

	Cluster1	Cluster2	Cluster3	Cluster4
Cluster1	0.0000	70.2162	85.685	69.834
Cluster2	70.2162	0.0000	84.138	98.975
Cluster3	85.6851	84.1377	0.000	133.935
Cluster4	69.8341	98.9752	133.935	0.000

Dendrogram



Resultado visualmente mais uniforme entre as observações dos quatro agrupamentos. Esta constatação é confirmada acima na quantidade de observações distribuídas da seguinte maneira: Cluster1 com 28 obs, Cluster2 com 34 obs, Cluster3 com 36 e Cluster4 com 34 obs.

Abaixo será executada a análise discriminante (Discriminant Analysis), e será mantida as variáveis que já vêm sendo utilizadas nos resultados acima ('Opportunity_n' 'Water withdrawals_np' 'Tolerance for immigrants_n' 'Religious tolerance_n' 'Governança_n' 'EPI-2014_n' 'HPI_n' 'Democracy Index-2014_n' 'IDH-2014_n').

Utilizar como preditora (Groups), a coluna C45, recém criada para armazenar os “Cluster memberships” e nomeada como “ClusterM”.

Resultados:

Discriminant Analysis: C45 versus Opportunity_n, Water withdrawals, ...

Linear Method for Response: C45

Predictors: Opportunity_n, Water withdrawals_np, Tolerance for immigrants_n,
Religious tolerance_n, Governança_n, EPI-2014_n, HPI_n,
Democracy Index-2014_n, IDH-2014_n

Group	1	2	3	4
Count	28	34	36	34

Summary of classification

Put into Group	True Group			
	1	2	3	4
1	27	0	0	0
2	1	32	2	0
3	0	2	34	0
4	0	0	0	34
Total N	28	34	36	34
N correct	27	32	34	34
Proportion	0.964	0.941	0.944	1.000

N = 132

N Correct = 127

Proportion Correct = 0.962

Squared Distance Between Groups

	1	2	3	4
1	0.0000	12.8992	27.5292	15.5903
2	12.8992	0.0000	20.3814	19.2151
3	27.5292	20.3814	0.0000	51.6822
4	15.5903	19.2151	51.6822	0.0000

Linear Discriminant Function for Groups

	1	2	3	4
Constant	-54.035	-33.655	-23.064	-66.764
Opportunity_n	-0.265	-0.299	-0.426	-0.139
Water withdrawals_np	0.087	0.064	0.121	0.100
Tolerance for immigrants_n	0.244	0.175	0.247	0.191
Religious tolerance_n	0.299	0.177	0.179	0.245
Governança_n	-0.336	-0.093	-0.090	-0.055
EPI-2014_n	0.262	0.227	0.045	0.268
HPI_n	0.163	0.133	0.116	0.121
Democracy Index-2014_n	0.438	0.231	0.322	0.297
IDH-2014_n	0.590	0.550	0.372	0.669

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
14**	3	2	1	21.86	0.004
			2	10.70	0.932
			3	16.07	0.064
			4	26.11	0.000
52**	2	3	1	38.00	0.000
			2	25.75	0.065
			3	20.41	0.935
			4	57.81	0.000
55**	3	2	1	27.83	0.001
			2	13.48	0.951
			3	19.45	0.048
			4	55.62	0.000
73**	1	2	1	10.006	0.469
			2	9.997	0.471
			3	31.029	0.000
			4	14.091	0.061
113**	2	3	1	26.81	0.001
			2	13.81	0.491
			3	13.74	0.508
			4	42.60	0.000

Nos resultados acima, algumas observações foram identificadas como não adequadas para seus agrupamentos, são elas a Bosnia and Herzegovina (linha 14) que está no grupo 3, e suggest to 2, a Índia (52) no 2, mas sugere no 3. O Iraq (55) no 3, mas sugere em 2, a Macedonia (73) no 1 e sugere no 2 e finalmente Swaziland (113) no 2, sugerindo no 3.

Dentre os grupos formados, os que estão mais próximos são os grupos 1 e 2, e os que estão mais distantes são os grupos 3 e 4.

Para o novo teste, foi alterado o “discriminant function” field de Linear para Quadratic, com os seguintes resultados:

Discriminant Analysis: ClusterM versus Opportunity, Water withdrawals, ...

Quadratic Method for Response: ClusterM

Predictors: Opportunity_n, Water withdrawals_np, Tolerance for immigrants_n, Religious tolerance_n, Governança_n, EPI-2014_n, HPI_n, Democracy Index-2014_n, IDH-2014_n

Group	1	2	3	4
Count	28	34	36	34

Summary of classification

Put into Group	True Group			
	1	2	3	4
1	28	0	0	0
2	0	34	1	0
3	0	0	35	0
4	0	0	0	34
Total N	28	34	36	34
N correct	28	34	35	34
Proportion	1.000	1.000	0.972	1.000

N = 132

N Correct = 131

Proportion Correct = 0.992

From Generalized Squared Distance to Group

Group	1	2	3	4
1	43.31	69.38	77.58	70.99
2	85.81	46.02	70.76	81.29
3	114.10	74.87	46.25	183.92
4	71.44	72.33	120.05	39.96

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
14**	3	2	1	110.70	0.000
			2	61.45	0.696
			3	63.11	0.304
			4	112.75	0.000

A quantidade de elementos corretos em seus devidos grupos (grau de acerto), aumentou de 127 para 131, given N 132.

A única observação que apresentou divergência de classificação foi a Bosnia and Herzegovina (linha 14) que está no grupo 3, e sugerido ir para 2. Portanto, a função discriminante quadrática na análise discriminante apresentou melhores resultados que a função discriminante Linear.

9.2. CONSIDERAÇÕES

O objetivo desta sessão foi realizar análises diversas a fim de reduzir a quantidade de variáveis do estudo, porém sem perder seu grau de explicação, significância nos resultados e a relevância do tema.

Seguindo a tendência das sessões anteriores, as variáveis que se destacaram, apresentaram um viés mais voltado ao capital social, que o natural. E mesmo reduzindo a quantidade para 8 (oito) variáveis, obtivemos um resultado satisfatório com a composição. O r-quadrado se manteve na ordem de 90% e o p-value < 0.05.

As observações foram distribuídas uniformemente nos 4 (quatro) agrupamentos, com alto grau de acerto (N 131). A única observação que apresentou divergência de classificação foi a Bosnia and Herzegovina.

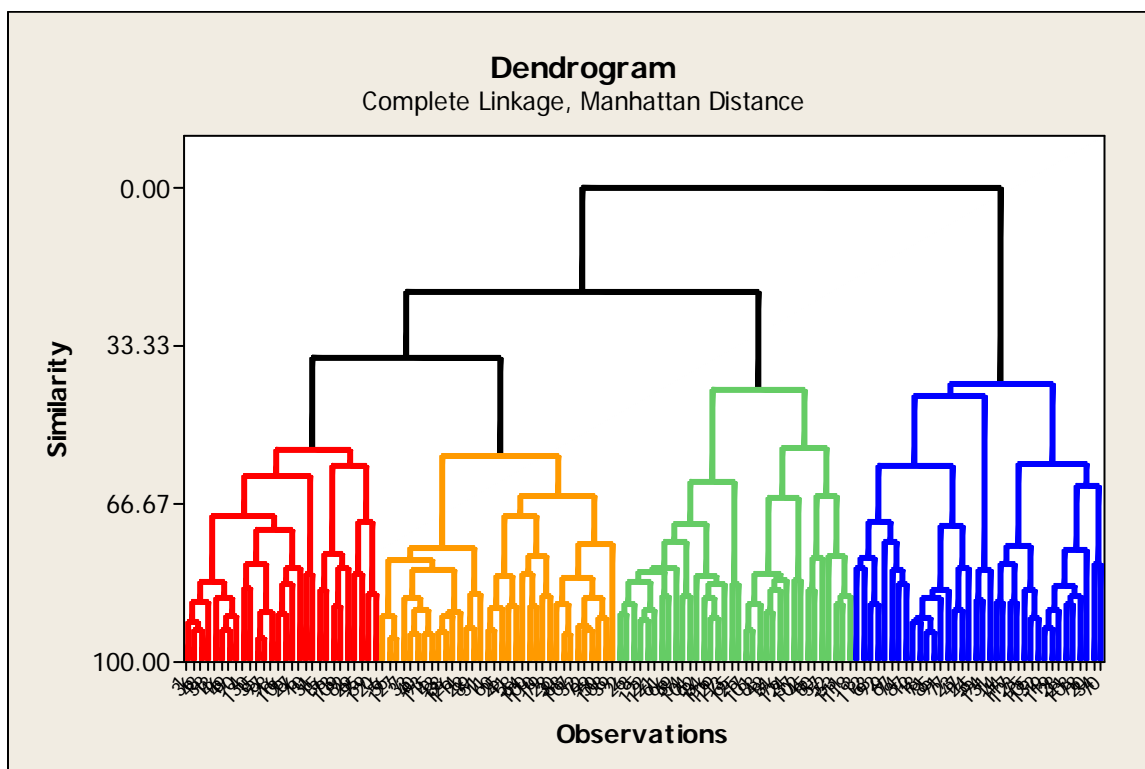
10. REGRESSÃO LOGÍSTICA

Serão executadas análises de regressão logística (*nominal e ordinal*) e discriminante (*linear e quadratic*) de um conjunto de dados para encontrar o modelo que melhor relacione as variáveis escolhidas, e com o maior o grau de acerto na distribuição das observações em seus respectivos grupos. Será mostrada uma tabela com os resultados das regressões logística e discriminante para efeito de comparação de efetividade dos modelos aqui analisados. Do total de 33, serão utilizadas apenas 9 (nove) variáveis previamente identificadas como as de melhor grau de explicação dos resultados.

10.1. Análises de Regressão Logística e Discriminante

Nesta sessão serão realizadas análises de regressão logística (*ordinal and nominal*) e discriminante (*linear and quadratic*), a fim de estabelecer o modelo que melhor represente os dados desta amostra.

As variáveis de trabalho para este estudo, são: EPI-2014, Governança, Democracy Index-2014, Opportunity, IDH-2014, Tolerance for immigrants, Water withdrawals, HPI (Happy Planet Index), Religious tolerance. Além destas 9 (nove) variáveis mencionadas, extraídas da base de dados deste estudo, 2 (duas) variáveis adicionais foram criadas para auxiliar nesta análise. A variável “**3Reg_reOrd**”, que representa as 3 (três) regiões AIBER, AVECO e OTHERS ordenadas pelos resultados de uma ANOVA (análise de variância), classificando os resultados dos mais baixos (piores) para os mais altos (melhores). E a variável “**ClusterM_reOrd**”, que é resultado de 4 (quatro) agrupamentos por similaridade, através do método *Complete* e medida de distância *Manhattan*, classificados em ordem crescente de média.



10.2 Análises de Regressão Logística (*nominal and ordinal*)

A análise de regressão logística tem como função evidenciar um modelo estatístico que melhor explique os dados e possa permitir predição de valores. É útil para modelar a probabilidade de um evento ocorrer como função de outros fatores. É comparável às técnicas supervisionadas utilizadas em aprendizagem automática, como árvores de decisão, redes neurais, etc... *fonte: Advanced Econometrics Harvard University Press via internet*

Neste trabalho, as análises de regressão logística terão como *response* as novas variáveis “ClusterM_reOrd” e “ClusterM_reOrd”, criadas para este estudo e como *predictors*, as 9 (nove) variáveis definidas na sessão anterior.

Iniciando com a regressão logística ordinal, temos:

Ordinal Logistic Regression: ClusterM_reOrd versus Opportunity, Water withdrawals,..

Link Function: Logit

Response Information

Variable	Value	Count
ClusterM_reOrd	1	36
	2	28
	3	34
	4	34
	Total	132

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper

Const(1)	3.72950	1.38597	2.69	0.007		
Const(2)	7.32211	1.55484	4.71	0.000		
Const(3)	11.2389	1.81172	6.20	0.000		
Opportunity_n	-0.0744932	0.0402693	-1.85	0.064	0.93	0.86
Water withdrawals_np	0.0123907	0.0079966	1.55	0.121	1.01	1.00
Tolerance for immigrants_n	0.0404440	0.0148498	2.72	0.006	1.04	1.01
Religious tolerance_n	0.0199327	0.0090083	2.21	0.027	1.02	1.00
Governança_n	-0.0887115	0.0269099	-3.30	0.001	0.92	0.87
EPI-2014_n	-0.0541366	0.0228767	-2.37	0.018	0.95	0.91
HPI_n	-0.0090037	0.0115265	-0.78	0.435	0.99	0.97
Democracy Index-2014_n	0.0667537	0.0229583	2.91	0.004	1.07	1.02
IDH-2014_n	-0.0586925	0.0270379	-2.17	0.030	0.94	0.89

Predictor	Upper
Const(1)	
Const(2)	
Const(3)	
Opportunity_n	1.00
Water withdrawals_np	1.03
Tolerance for immigrants_n	1.07
Religious tolerance_n	1.04
Governança_n	0.96
EPI-2014_n	0.99
HPI_n	1.01
Democracy Index-2014_n	1.12
IDH-2014_n	0.99

Log-Likelihood = -71.351

Test that all slopes are zero: G = 222.156, DF = 9, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	218.385	384	1.000
Deviance	142.702	384	1.000

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	6209	95.3	Somers' D 0.91
Discordant	301	4.6	Goodman-Kruskal Gamma 0.91
Ties	6	0.1	Kendall's Tau-a 0.68
Total	6516	100.0	

Apesar do nível de concordância chegar a 95.3%, duas variáveis apresentaram p-value acima do esperado e serão eliminadas da análise, são elas:

Water withdrawals_np	p-value = 0.121
HPI_n	p-value = 0.435

Após executado novamente, seguem os resultados:

Ordinal Logistic Regression: ClusterM_reOrd versus Opportunity, Tolerance, ..

Link Function: Logit

Response Information

Variable	Value	Count
ClusterM_reOrd	1	36
	2	28
	3	34
	4	34

Total 132

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Const(1)	3.94355	1.22689	3.21	0.001			
Const(2)	7.29842	1.41989	5.14	0.000			
Const(3)	11.2859	1.68620	6.69	0.000			
Opportunity_n	-0.0798587	0.0401193	-1.99	0.047	0.92	0.85	
Tolerance for immigrants_n	0.0389503	0.0145887	2.67	0.008	1.04	1.01	
Religious tolerance_n	0.0207656	0.0088759	2.34	0.019	1.02	1.00	
Governança_n	-0.0783726	0.0247985	-3.16	0.002	0.92	0.88	
EPI-2014_n	-0.0442065	0.0217034	-2.04	0.042	0.96	0.92	
Democracy Index-2014_n	0.0683465	0.0220906	3.09	0.002	1.07	1.03	
IDH-2014_n	-0.0707820	0.0261333	-2.71	0.007	0.93	0.89	

Predictor	Upper
Const(1)	
Const(2)	
Const(3)	
Opportunity_n	1.00
Tolerance for immigrants_n	1.07
Religious tolerance_n	1.04
Governança_n	0.97
EPI-2014_n	1.00
Democracy Index-2014_n	1.12
IDH-2014_n	0.98

Log-Likelihood = -72.832

Test that all slopes are zero: G = 219.194, DF = 7, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	237.835	386	1.000
Deviance	145.664	386	1.000

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	6184	94.9	Somers' D 0.90
Discordant	328	5.0	Goodman-Kruskal Gamma 0.90
Ties	4	0.1	Kendall's Tau-a 0.68
Total	6516	100.0	

Todos com p-value dentro da faixa esperada e manteve-se o nível de concordância próximo ao resultado anterior, ou seja em 94.9%.

Agora executando novamente, mas tendo a coluna das 3 regiões ordenadas (3Reg_reOrd) como *response* e todas as variáveis, temos:

Ordinal Logistic Regression: 3Reg_reOrd versus Opportunity, Water withdrawals, ...

Link Function: Logit

Response Information

Variable	Value	Count
3Reg_reOrd	1	84
	2	21
	3	27
	Total	132

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Const(1)	19.0782	3.72232	5.13	0.000			
Const(2)	21.8214	3.97668	5.49	0.000			
Opportunity_n	-0.0636396	0.0562010	-1.13	0.257	0.94	0.84	
Water withdrawals_np	-0.0180067	0.0110094	-1.64	0.102	0.98	0.96	
Tolerance for immigrants_n	0.0104075	0.0185100	0.56	0.574	1.01	0.97	
Religious tolerance_n	-0.0241964	0.0117145	-2.07	0.039	0.98	0.95	
Governança_n	0.0364954	0.0386019	0.95	0.344	1.04	0.96	
EPI-2014_n	0.0205287	0.0357756	0.57	0.566	1.02	0.95	
HPI_n	-0.0565586	0.0174538	-3.24	0.001	0.95	0.91	
Democracy Index-2014_n	-0.0505792	0.0438318	-1.15	0.249	0.95	0.87	
IDH-2014_n	-0.153374	0.0515304	-2.98	0.003	0.86	0.78	

Predictor	Upper
Const(1)	
Const(2)	
Opportunity_n	1.05
Water withdrawals_np	1.00
Tolerance for immigrants_n	1.05
Religious tolerance_n	1.00
Governança_n	1.12
EPI-2014_n	1.09
HPI_n	0.98
Democracy Index-2014_n	1.04
IDH-2014_n	0.95

Log-Likelihood = -47.494

Test that all slopes are zero: G = 143.849, DF = 9, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	101.921	253	1.000
Deviance	94.988	253	1.000

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	4414	96.0	Somers' D 0.92
Discordant	176	3.8	Goodman-Kruskal Gamma 0.92
Ties	9	0.2	Kendall's Tau-a 0.49
Total	4599	100.0	

As seguintes variáveis apresentaram p-value acima do esperado e serão retiradas para realizar nova análise:

Opportunity_n	p-value = 0.257
Tolerance for immigrants_n	p-value = 0.574
Governança_n	p-value = 0.344
EPI-2014_n	p-value = 0.566
Democracy Index-2014_n	p-value = 0.249

Por questão de parcimônia, a variável “**water withdrawals_np**”, que apresentou p-value de 0.102, será mantida para a análise que segue.

Ordinal Logistic Regression: 3Reg_reOrd versus Water withdrawals, Religious, ...

Link Function: Logit

Response Information

Variable	Value	Count
3Reg_reOrd	1	84
	2	21
	3	27
	Total	132

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Const(1)	20.2490	3.22592	6.28	0.000			
Const(2)	22.6211	3.47550	6.51	0.000			
Water withdrawals_np	-0.0223421	0.0096957	-2.30	0.021	0.98	0.96	1.00
Religious tolerance_n	-0.0327494	0.0098736	-3.32	0.001	0.97	0.95	0.99
HPI_n	-0.0653890	0.0170031	-3.85	0.000	0.94	0.91	0.97
IDH-2014_n	-0.185861	0.0289818	-6.41	0.000	0.83	0.78	0.88

Log-Likelihood = -52.948

Test that all slopes are zero: G = 132.942, DF = 4, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	115.006	258	1.000
Deviance	105.895	258	1.000

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	4378	95.2	Somers' D 0.90
Discordant	217	4.7	Goodman-Kruskal Gamma 0.91
Ties	4	0.1	Kendall's Tau-a 0.48
Total	4599	100.0	

A quantidade de variáveis foi reduzida para 4 (quatro), todas apresentaram um p-value dentro da faixa esperada, inclusive a variável “water withdrawals_np”. O nível de concordância ficou em 95.2% contra os 96.0% no resultado anterior quando executado com as 9 (nove) variáveis.

Abaixo iremos executar a função “nominal logistic regression” com a nova coluna dos clusters ordenados (ClusterM_reOrd) e com a nova coluna das regiões ordenadas (3Reg_reOrd), como variáveis *response*.

Nominal Logistic Regression: ClusterM_reOrd versus Opportunity, Water withdrawals,

```
* WARNING * Algorithm has not converged after 20 iterations.
* WARNING * Convergence has not been reached for the parameter estimates
            criterion.
* WARNING * The results may not be reliable.
* WARNING * Try increasing the maximum number of iterations.
```

Response Information

Variable	Value	Count
ClusterM_reOrd	4	34 (Reference Event)
	3	34
	2	28
	1	36
	Total	132

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	CI Lower
Logit 1: (3/4)						
Constant	283.644	26330.3	0.01	0.991		
Opportunity_n	-0.739784	472.076	-0.00	0.999	0.48	0.00
Water withdrawals_np	-0.457341	73.6768	-0.01	0.995	0.63	0.00
Tolerance for immigrants_n	-0.646618	173.300	-0.00	0.997	0.52	0.00
Religious tolerance_n	-1.06655	161.395	-0.01	0.995	0.34	0.00
Governança_n	1.05558	407.672	0.00	0.998	2.87	0.00
EPI-2014_n	-0.0587769	302.559	-0.00	1.000	0.94	0.00
HPI_n	0.230493	113.474	0.00	0.998	1.26	0.00
Democracy Index-2014_n	-1.00389	332.920	-0.00	0.998	0.37	0.00
IDH-2014_n	-1.72852	496.859	-0.00	0.997	0.18	0.00
Logit 2: (2/4)						
Constant	78.2069	44733.8	0.00	0.999		
Opportunity_n	-1.04789	575.002	-0.00	0.999	0.35	0.00
Water withdrawals_np	-0.126672	66.9063	-0.00	0.998	0.88	0.00
Tolerance for immigrants_n	0.451843	203.967	0.00	0.998	1.57	0.00
Religious tolerance_n	0.495841	235.964	0.00	0.998	1.64	0.00
Governança_n	-1.88784	442.534	-0.00	0.997	0.15	0.00
EPI-2014_n	0.856805	205.541	0.00	0.997	2.36	0.00
HPI_n	0.0711650	58.2793	0.00	0.999	1.07	0.00
Democracy Index-2014_n	1.21463	526.663	0.00	0.998	3.37	0.00
IDH-2014_n	-1.59111	618.811	-0.00	0.998	0.20	0.00
Logit 3: (1/4)						
Constant	344.971	28900.9	0.01	0.990		
Opportunity_n	-4.03676	641.182	-0.01	0.995	0.02	0.00
Water withdrawals_np	0.493440	90.1376	0.01	0.996	1.64	0.00
Tolerance for immigrants_n	0.317522	210.905	0.00	0.999	1.37	0.00
Religious tolerance_n	-0.100582	224.256	-0.00	1.000	0.90	0.00
Governança_n	-0.536349	515.493	-0.00	0.999	0.58	0.00
EPI-2014_n	-2.02539	333.734	-0.01	0.995	0.13	0.00
HPI_n	-0.536653	175.247	-0.00	0.998	0.58	0.00
Democracy Index-2014_n	0.294046	313.633	0.00	0.999	1.34	0.00
IDH-2014_n	-0.974498	506.512	-0.00	0.998	0.38	0.00

Predictor	Upper
Logit 1: (3/4)	
Constant	
Opportunity_n	*
Water withdrawals_np	3.27456E+62
Tolerance for immigrants_n	1.70687E+147
Religious tolerance_n	8.25362E+136
Governança_n	*
EPI-2014_n	3.25846E+257
HPI_n	4.89379E+96
Democracy Index-2014_n	8.83108E+282
IDH-2014_n	*
Logit 2: (2/4)	
Constant	
Opportunity_n	*
Water withdrawals_np	7.86551E+56
Tolerance for immigrants_n	6.50543E+173
Religious tolerance_n	1.17005E+201
Governança_n	*
EPI-2014_n	2.13108E+175
HPI_n	4.34834E+49
Democracy Index-2014_n	*
IDH-2014_n	*
Logit 3: (1/4)	
Constant	
Opportunity_n	*
Water withdrawals_np	8.70133E+76
Tolerance for immigrants_n	4.57446E+179
Religious tolerance_n	6.96726E+190
Governança_n	*
EPI-2014_n	1.56828E+283
HPI_n	8.65975E+148
Democracy Index-2014_n	1.23780E+267
IDH-2014_n	*

Log-Likelihood = -0.000
 Test that all slopes are zero: G = 364.858, DF = 27, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	0.0000018	366	1.000
Deviance	0.0000036	366	1.000

O método de regressão logística *Nominal* não conseguiu analisar adequadamente os dados de nossa base.

Tomando a coluna das 3 (três) regiões ordenadas, e executando a função *nominal logistic regression*, temos os seguintes resultados

Nominal Logistic Regression: 3Reg_reOrd versus Opportunity, Water withdrawals, ...

* WARNING * Algorithm has not converged after 20 iterations.
 * WARNING * Convergence has not been reached for the parameter estimates criterion.
 * WARNING * The results may not be reliable.
 * WARNING * Try increasing the maximum number of iterations.

Response Information

Variable	Value	Count	
3Reg_reOrd	3	27	(Reference Event)
	2	21	
	1	84	
	Total	132	

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower
Logit 1: (2/3)						
Constant	1787.37	26631.6	0.07	0.946		
Opportunity_n	5.60972	462.674	0.01	0.990	273.07	0.00
Water withdrawals_np	-0.411068	31.4881	-0.01	0.990	0.66	0.00
Tolerance for immigrants_n	0.968739	97.7937	0.01	0.992	2.63	0.00
Religious tolerance_n	-1.67529	75.8557	-0.02	0.982	0.19	0.00
Governança_n	0.252903	394.961	0.00	0.999	1.29	0.00
EPI-2014_n	1.59830	274.719	0.01	0.995	4.94	0.00
HPI_n	-1.66043	172.874	-0.01	0.992	0.19	0.00
Democracy Index-2014_n	-12.3069	1412.70	-0.01	0.993	0.00	0.00
IDH-2014_n	-14.7970	1014.67	-0.01	0.988	0.00	0.00
Logit 2: (1/3)						
Constant	1813.46	26631.6	0.07	0.946		
Opportunity_n	5.51325	462.674	0.01	0.990	247.96	0.00
Water withdrawals_np	-0.448119	31.4881	-0.01	0.989	0.64	0.00
Tolerance for immigrants_n	0.914868	97.7937	0.01	0.993	2.50	0.00
Religious tolerance_n	-1.71150	75.8557	-0.02	0.982	0.18	0.00
Governança_n	0.368590	394.961	0.00	0.999	1.45	0.00
EPI-2014_n	1.62405	274.719	0.01	0.995	5.07	0.00
HPI_n	-1.77813	172.874	-0.01	0.992	0.17	0.00
Democracy Index-2014_n	-12.3383	1412.70	-0.01	0.993	0.00	0.00
IDH-2014_n	-14.9646	1014.67	-0.01	0.988	0.00	0.00

Predictor Upper

Logit 1: (2/3)	
Constant	
Opportunity_n	*
Water withdrawals_np	4.20912E+26
Tolerance for immigrants_n	4.60105E+83

```

Religious tolerance_n      6.93363E+63
Governança_n              *
EPI-2014_n                3.43044E+234
HPI_n                     2.68669E+146
Democracy Index-2014_n   *
IDH-2014_n                *
Logit 2: (1/3)
Constant
Opportunity_n              *
Water withdrawals_np      4.05603E+26
Tolerance for immigrants_n 4.35970E+83
Religious tolerance_n     6.68700E+63
Governança_n              *
EPI-2014_n                3.51992E+234
HPI_n                     2.38838E+146
Democracy Index-2014_n   *
IDH-2014_n                *

```

Log-Likelihood = -13.036

Test that all slopes are zero: G = 212.766, DF = 18, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	43.6505	244	1.000
Deviance	26.0716	244	1.000

O método de regressão logística *Nominal* não conseguiu analisar adequadamente os dados de nossa base.

10.3 Análise Discriminante (*linear and quadratic*)

Nesta sessão será executada a função de Análise Discriminante (*discriminant analysis*), tendo como *response* primeiramente a coluna “ClusterM_reOrd” e na sequência a coluna “3Reg_reOrd”. As variáveis preditoras serão as mesmas utilizadas nas regressões da sessão anterior, ou seja, 'Opportunity_n' 'Water withdrawals_np' 'Tolerance for immigrants_n' 'Religious tolerance_n' 'Governança_n' 'EPI-2014_n' 'HPI_n' 'Democracy Index-2014_n' 'IDH-2014_n'.

A análise discriminante é uma técnica da estatística multivariada utilizada para discriminar e classificar objetos. Segundo KHATTREE & NAIK (2000) é uma técnica da estatística multivariada que estuda a separação de objetos de uma população em duas ou mais classes. *fonte: apostila ANÁLISE MULTIVARIADA APLICADA AS CIÊNCIAS AGRÁRIAS - PÓS-GRADUAÇÃO EM AGRONOMIA CIÊNCIA DO SOLO: CPGA – Prof. Carlos Alberto Alves Varella, da Universidade Federal Rural do Rio de Janeiro.*

Iniciamos com o método linear da análise discriminante, cujos resultados são:

Discriminant Analysis: ClusterM_reOrd versus Opportunity, Water withdrawals, ...

Linear Method for Response: ClusterM_reOrd

Predictors: Opportunity_n, Water withdrawals_np, Tolerance for immigrants_n,
 Religious tolerance_n, Governança_n, EPI-2014_n, HPI_n,
 Democracy Index-2014_n, IDH-2014_n

Group	1	2	3	4
-------	---	---	---	---

Count 36 28 34 34

Summary of classification

Put into Group	True Group			
	1	2	3	4
1	34	0	2	0
2	0	27	0	0
3	2	1	32	0
4	0	0	0	34
Total N	36	28	34	34
N correct	34	27	32	34
Proportion	0.944	0.964	0.941	1.000

N = 132

N Correct = 127

Proportion Correct = 0.962

Squared Distance Between Groups

	1	2	3	4
1	0.0000	27.5292	20.3814	51.6822
2	27.5292	0.0000	12.8992	15.5903
3	20.3814	12.8992	0.0000	19.2151
4	51.6822	15.5903	19.2151	0.0000

Linear Discriminant Function for Groups

	1	2	3	4
Constant	-23.064	-54.035	-33.655	-66.764
Opportunity_n	-0.426	-0.265	-0.299	-0.139
Water withdrawals_np	0.121	0.087	0.064	0.100
Tolerance for immigrants_n	0.247	0.244	0.175	0.191
Religious tolerance_n	0.179	0.299	0.177	0.245
Governança_n	-0.090	-0.336	-0.093	-0.055
EPI-2014_n	0.045	0.262	0.227	0.268
HPI_n	0.116	0.163	0.133	0.121
Democracy Index-2014_n	0.322	0.438	0.231	0.297
IDH-2014_n	0.372	0.590	0.550	0.669

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
14**	1	3	1	16.07	0.064
			2	21.86	0.004
			3	10.70	0.932
			4	26.11	0.000
52**	3	1	1	20.41	0.935
			2	38.00	0.000
			3	25.75	0.065
55**	1	3	4	57.81	0.000
			1	19.45	0.048
			2	27.83	0.001
			3	13.48	0.951
73**	2	3	4	55.62	0.000
			1	31.029	0.000
			2	10.006	0.469
			3	9.997	0.471
113**	3	1	4	14.091	0.061
			1	13.74	0.508
			2	26.81	0.001
			3	13.81	0.491
			4	42.60	0.000

Em seguida, passamos para o método quadrático da análise discriminante, para as variáveis ClusterM_reOrd e 3Reg_reOrd, cujos resultados são:

Discriminant Analysis: ClusterM_reOrd versus Opportunity, Water withdrawals, ...

Quadratic Method for Response: ClusterM_reOrd

Predictors: Opportunity_n, Water withdrawals_np, Tolerance for immigrants_n,
Religious tolerance_n, Governança_n, EPI-2014_n, HPI_n,
Democracy Index-2014_n, IDH-2014_n

Group	1	2	3	4
Count	36	28	34	34

Summary of classification

Put into Group	True Group			
	1	2	3	4
1	35	0	0	0
2	0	28	0	0
3	1	0	34	0
4	0	0	0	34
Total N	36	28	34	34
N correct	35	28	34	34
Proportion	0.972	1.000	1.000	1.000

N = 132

N Correct = 131

Proportion Correct = 0.992

From Generalized Squared Distance to Group

Group	1	2	3	4
1	46.25	114.10	74.87	183.92
2	77.58	43.31	69.38	70.99
3	70.76	85.81	46.02	81.29
4	120.05	71.44	72.33	39.96

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
14**	1	3	1	63.11	0.304
			2	110.70	0.000
			3	61.45	0.696
			4	112.75	0.000

A análise discriminante utilizando o método *Linear*, apresentou bons resultados com 96.2% de acertos, e 5 (cinco) países mal classificados do total de 132. Já no método *Quadratic*, a proporção de resultados corretos passou a 99.2% , apresentando apenas 1 país como mal classificado. Portanto, vamos utilizar os resultados do método *Quadratic* para nossas considerações.

A observação na linha 14 refere-se o país Bosnia and Herzegovina, que inicialmente foi categorizado como 1 e pela análise do software estatístico, deveria ser realocado no grupo 3.

Executando a análise tendo a coluna '3Reg_reOrd' como *response*, obtivemos os seguintes resultados

Discriminant Analysis: 3Reg_reOrd versus Opportunity, Water withdrawals, ...

Linear Method for Response: 3Reg_reOrd

Predictors: Opportunity_n, Water withdrawals_np, Tolerance for immigrants_n,
Religious tolerance_n, Governança_n, EPI-2014_n, HPI_n,
Democracy Index-2014_n, IDH-2014_n

Group	1	2	3
-------	---	---	---

Count 84 21 27

Summary of classification

Put into Group	True Group		
	1	2	3
1	75	1	2
2	5	16	0
3	4	4	25
Total N	84	21	27
N correct	75	16	25
Proportion	0.893	0.762	0.926

N = 132

N Correct = 116

Proportion Correct = 0.879

Squared Distance Between Groups

	1	2	3
1	0.0000	9.5087	11.5092
2	9.5087	0.0000	11.3303
3	11.5092	11.3303	0.0000

Linear Discriminant Function for Groups

	1	2	3
Constant	-22.171	-42.275	-39.069
Opportunity_n	-0.415	-0.328	-0.291
Water withdrawals_np	0.117	0.139	0.145
Tolerance for immigrants_n	0.251	0.315	0.263
Religious tolerance_n	0.116	0.165	0.110
Governança_n	-0.059	-0.221	0.021
EPI-2014_n	0.013	0.032	-0.027
HPI_n	0.196	0.315	0.229
Democracy Index-2014_n	0.221	0.285	0.210
IDH-2014_n	0.367	0.424	0.441

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
1**	1	2	1	9.652	0.040
			2	3.291	0.957
			3	14.583	0.003
25**	2	3	1	14.361	0.007
			2	11.650	0.028
			3	4.547	0.965
45**	3	1	1	6.440	0.771
			2	14.386	0.015
			3	9.000	0.214
48**	1	2	1	8.973	0.040
			2	2.628	0.959
			3	16.915	0.001
50**	1	3	1	7.248	0.185
			2	12.287	0.015
			3	4.315	0.800
59**	1	2	1	23.790	0.001
			2	9.238	0.999
			3	26.074	0.000
68**	3	1	1	13.04	0.498
			2	20.28	0.013
			3	13.07	0.489
72**	1	3	1	9.885	0.343
			2	21.011	0.001
			3	8.590	0.656
79**	1	3	1	10.830	0.170
			2	14.678	0.025
			3	7.715	0.806
80**	2	1	1	7.230	0.542
			2	7.921	0.383

			3	11.187	0.075
98**	1	2	1	12.975	0.028
			2	5.886	0.970
			3	17.925	0.002
99**	1	3	1	9.932	0.071
			2	15.201	0.005
			3	4.802	0.924
100**	2	3	1	12.507	0.057
			2	10.594	0.149
			3	7.253	0.793
106**	1	2	1	10.124	0.256
			2	8.077	0.712
			3	14.274	0.032
110**	2	3	1	13.657	0.021
			2	8.447	0.284
			3	6.654	0.695
128**	2	3	1	14.037	0.010
			2	7.623	0.246
			3	5.405	0.744

Executando o método: *Quadratic*

Discriminant Analysis: 3Reg_reOrd versus Opportunity, Water withdrawals, ...

Quadratic Method for Response: 3Reg_reOrd

Predictors: Opportunity_n, Water withdrawals_np, Tolerance for immigrants_n,
Religious tolerance_n, Governança_n, EPI-2014_n, HPI_n,
Democracy Index-2014_n, IDH-2014_n

Group	1	2	3
Count	84	21	27

Summary of classification

Put into Group	True Group		
	1	2	3
1	78	0	0
2	2	21	0
3	4	0	27
Total N	84	21	27
N correct	78	21	27
Proportion	0.929	1.000	1.000

N = 132

N Correct = 126

Proportion Correct = 0.955

From Generalized Squared Distance to Group

Group	1	2	3
1	50.39	73.20	425.97
2	59.05	42.27	183.60
3	63.63	60.17	37.67

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
1**	1	2	1	58.70	0.004
			2	47.65	0.996
			3	164.95	0.000
30**	1	3	1	58.89	0.024
			2	121.98	0.000
			3	51.50	0.976
48**	1	2	1	57.85	0.009
			2	48.35	0.991
			3	394.52	0.000

50**	1	3	1	56.56	0.147
			2	78.84	0.000
			3	53.03	0.853
72**	1	3	1	58.43	0.001
			2	106.85	0.000
			3	44.86	0.999
99**	1	3	1	58.68	0.048
			2	76.39	0.000
			3	52.70	0.952

No método *linear*, obtivemos 87.9% de acerto, resultando em 16 erros, ou 16 países mal alocados conforme análise do *software*. No método *quadratic* a quantidade de erros caiu para 6 e a proporção de acertos subiu para 95.5%, portanto este será o método escolhido para este estudo.

Os 6 países que figuraram como mal classificados pela ferramenta, pertencem ao agrupamento OTHERS, são eles Albânia (obs 1), Croácia (obs 30), Guyana (obs 48), Hungria (obs 50), Lituânia (obs 72) e Filipinas (obs 99).

10.4 Resultados das análises Logística e Discriminante

Nesta sessão, apresentamos uma tabela contendo o conjunto dos resultados dos testes de modelos realizados.

Modelo	ClusterM_reOrd	3Reg_reOrd
AD (ANÁLISE DISCRIMINANTE)	99.2% (Q)	95.5% (Q)
RL (REGRESSÃO LOGÍSTICA)	94.9% (O)	95.2% (O)

legenda:

ClusterM_reOrd = 4 clusters ordenados, sendo de 1 (pior) a 4 (melhor)

3Reg_reOrd = 3 regiões ordenadas, sendo 1 para Others, 2 AIBER e 3 para AVECO

Q = método *quadratic* na análise discriminante

O = método *ordinal* na regressão logística

Na tabela vemos que o modelo de análise discriminante apresentou melhores resultados que o de regressão logística.

Apesar dos resultados se apresentarem relativamente próximos, cabe ressaltar que no modelo de análise discriminante (AD), foi possível trabalhar com todas as variáveis em todos os métodos, já para a regressão logística (RL), foram consideradas como relevantes 7 (sete) variáveis quando analisando a ClusterM_reOrd, e 5 (cinco) para a variável 3Reg_reOrd, mostrando portanto robustez superior no primeiro modelo.

10.5. CONSIDERAÇÕES

Como resultado das análises efetuadas, vemos na tabela da sessão 10.4, que o modelo de análise discriminante quadrática foi o que se apresentou mais robustamente para classificar e discriminar os objetos deste estudo.

Em relação ao tema Biodiversidade – Capital Natural e Social, objetivo central desta análise, percebemos nos últimos estudos que as variáveis de maior significância apresentam um viés mais voltado ao capital social que o natural. Talvez a mensagem que está tentando emergir na evolução deste estudo é a de que o indivíduo, assim como na pirâmide de Maslow, necessita primeiramente tratar dos temas que o impacta diretamente (comunidade local, respeito e oportunidades), para depois se dedicar ao entorno (proteção e preservação dos ativos naturais, qualidade de vida e humanidade).

11. ÁRVORES DE CLASSIFICAÇÃO

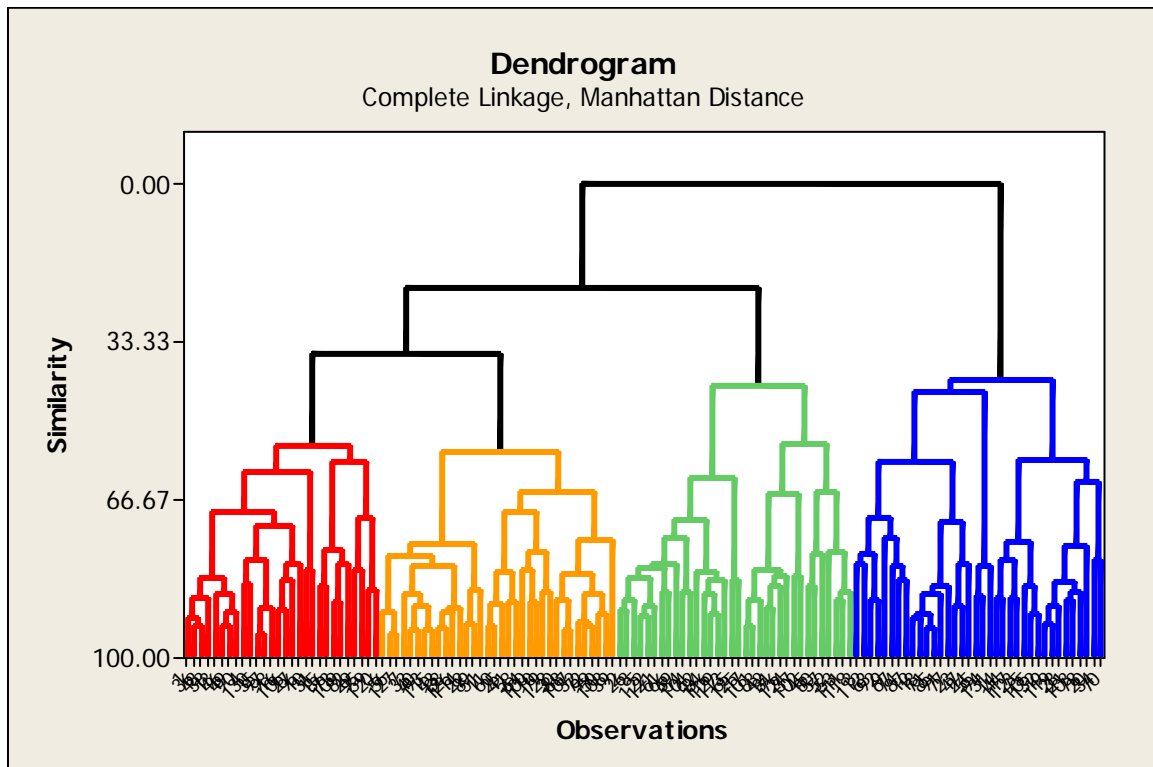
Nesta sessão serão realizadas análises discriminante e de árvores de classificação (árvores de decisão), para classificar hierarquicamente os dados via teste lógico da ferramenta SPSS e encontrar o modelo que melhor relacione as variáveis escolhidas. Ao fim será mostrada uma tabela com os resultados da análises discriminantes efetuadas nos *softwares* Minitab e SPSS para fins comparativos de efetividade.

11.1. Tratamento dos dados, análises discriminante e de árvores de classificação

Os dados originalmente armazenados na base do Minitab, foram exportados para o MS-Excel, tratados e importados no *software* SPSS para a realização das análises constantes neste estudo.

Nesta sessão serão realizadas análises discriminante e de árvores de classificação (árvores de decisão), a fim de estabelecer o modelo que melhor represente os dados desta amostra.

As variáveis de trabalho para este estudo, conforme detalhadas na sessão anterior (quadro 1), são: EPI-2014, Governança, Democracy Index-2014, Opportunity, IDH-2014, Tolerance for immigrants, Water withdrawals, HPI (Happy Planet Index), Religious tolerance. Além destas 9 (nove) variáveis mencionadas, extraídas da base de dados deste estudo, 2 (duas) variáveis adicionais foram criadas para auxiliar nesta análise. A variável “**3Reg_reOrd**”, que representa as 3 (três) regiões AIBER, AVECO e OTHERS ordenadas pelos resultados de uma ANOVA (análise de variância), classificando os resultados dos mais baixos (piores) para os mais altos (melhores). E a variável “**ClusterM_reOrd**”, que é resultado de 4 (quatro) agrupamentos por similaridade, através do método *Complete* e medida de distância *Manhattan*, classificados em ordem crescente de média.



11.2 Análise Discriminante (utilizando o software SPSS)

Nesta sessão será executada a função de Análise Discriminante (*discriminant analysis*), tendo como *response (dependent variable)* primeiramente a coluna “ClusterM_reOrd” e na sequência a coluna “3Reg_reOrd”. As variáveis predictoras (*independent variables*) serão: 'Opportunity_n' 'Water withdrawals_np' 'Tolerance for immigrants_n' 'Religious tolerance_n' 'Governança_n' 'EPI-2014_n' 'HPI_n' 'Democracy Index-2014_n' 'IDH-2014_n'.

A análise discriminante é uma técnica da estatística multivariada utilizada para discriminar e classificar objetos. Segundo KHATTREE & NAIK (2000) é uma técnica da estatística multivariada que estuda a separação de objetos de uma população em duas ou mais classes.

fonte: apostila ANÁLISE MULTIVARIADA APLICADA AS CIÊNCIAS AGRÁRIAS - PÓS-GRADUAÇÃO EM AGRONOMIA CIÊNCIA DO SOLO: CPGA – Prof. Carlos Alberto Alves Varella, da Universidade Federal Rural do Rio de Janeiro.

RESULTADOS: Análise Discriminante: by GROUPS=ClusterM_reOrd

GET

FILE='C:\Users\Arvores de Classificacao.sav'.

DATASET NAME DataSet1 WINDOW=FRONT.

DISCRIMINANT

/GROUPS=ClusterM_reOrd(1 3)

/VARIABLES=Opportunity_n Waterwithdrawals_np Toleranceforimmigrants_n Religiousolerance_n IDH2014_n Governança_n EPI2014_n HPI_n DemocracyIndex2014_n

/ANALYSIS ALL

```

/METHOD=WILKS
/FIN=3.84
/FOUT=2.71
/PRIORS EQUAL
/HISTORY
/STATISTICS=TABLE
/CLASSIFY=NONMISSING POOLED.

```

Discriminant

Notes

Output Created		22-May-2016 15:24:13
Comments		
Input	Data	C:\Users\Arvores de Classificacao.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data	132
	File	
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing in the analysis phase.
	Cases Used	In the analysis phase, cases with no user- or system-missing values for any predictor variable are used. Cases with user-, system-missing, or out-of-range values for the grouping variable are always excluded.
Syntax		DISCRIMINANT /GROUPS=ClusterM_reOrd(1 3) /VARIABLES=Opportunity_n Waterwithdrawals_np Toleranceforimmigrants_n Religioustolerance_n IDH2014_n Governana_n EPI2014_n HPI_n DemocracyIndex2014_n /ANALYSIS ALL /METHOD=WILKS /FIN=3.84 /FOUT=2.71 /PRIORS EQUAL /HISTORY /STATISTICS=TABLE /CLASSIFY=NONMISSING POOLED.
Resources	Processor Time	00:00:00.047

Notes

Output Created		22-May-2016 15:24:13
Comments		
Input	Data	C:\Users\Arvores de Classificacao.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data	132
	File	
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing in the analysis phase.
	Cases Used	In the analysis phase, cases with no user- or system-missing values for any predictor variable are used. Cases with user-, system-missing, or out-of-range values for the grouping variable are always excluded.
Syntax		<p>DISCRIMINANT</p> <p>/GROUPS=ClusterM_reOrd(1 3)</p> <p>/VARIABLES=Opportunity_n Waterwithdrawals_np Toleranceforimmigrants_n Religioustolerance_n IDH2014_n Governança_n EPI2014_n HPI_n DemocracyIndex2014_n</p> <p>/ANALYSIS ALL</p> <p>/METHOD=WILKS</p> <p>/FIN=3.84</p> <p>/FOUT=2.71</p> <p>/PRIORS EQUAL</p> <p>/HISTORY</p> <p>/STATISTICS=TABLE</p> <p>/CLASSIFY=NONMISSING POOLED.</p>
Resources	Processor Time	00:00:00.047
	Elapsed Time	00:00:00.077

[DataSet1] C:\Users\Arvores de Classificacao.sav

Analysis Case Processing Summary

Unweighted Cases	N	Percent
Valid	98	74.2

Excluded	Missing or out-of-range group codes	34	25.8
	At least one missing discriminating variable	0	.0
	Both missing or out-of-range group codes and at least one missing discriminating variable	0	.0
	Total	34	25.8
Total		132	100.0

Group Statistics

ClusterM_reOrd		Valid N (listwise)	
		Unweighted	Weighted
1	Opportunity	36	36.000
	Water withdrawals_np	36	36.000
	Tolerance for immigrants_n	36	36.000
	Religious tolerance_n	36	36.000
	IDH-2014_n	36	36.000
	Governance	36	36.000
	EPI-2014_n	36	36.000
	HPI	36	36.000
	Democracy Index-2014_n	36	36.000
2	Opportunity	28	28.000
	Water withdrawals_np	28	28.000
	Tolerance for immigrants_n	28	28.000
	Religious tolerance_n	28	28.000
	IDH-2014_n	28	28.000
	Governance	28	28.000
	EPI-2014_n	28	28.000
	HPI	28	28.000
	Democracy Index-2014_n	28	28.000
3	Opportunity	34	34.000
	Water withdrawals_np	34	34.000
	Tolerance for immigrants_n	34	34.000
	Religious tolerance_n	34	34.000
	IDH-2014_n	34	34.000
	Governance	34	34.000
	EPI-2014_n	34	34.000
	HPI	34	34.000

	Democracy Index-2014_n	34	34.000
Total	Opportunity	98	98.000
	Water withdrawals_np	98	98.000
	Tolerance for immigrants_n	98	98.000
	Religious tolerance_n	98	98.000
	IDH-2014_n	98	98.000
	Governance	98	98.000
	EPI-2014_n	98	98.000
	HPI	98	98.000
	Democracy Index-2014_n	98	98.000

Analysis 1

Stepwise Statistics

Variables Entered/Removed^{a,b,c,d}

Step	Entered	Wilks' Lambda			
		Statistic	df1	df2	df3
1	IDH-2014_n	.285	1	2	95.000
2	Religious tolerance_n	.158	2	2	95.000
3	Democracy Index-2014_n	.128	3	2	95.000
4	Governance	.101	4	2	95.000
5	EPI-2014_n	.087	5	2	95.000
6	Tolerance for immigrants_n	.078	6	2	95.000
7	Water withdrawals_np	.071	7	2	95.000

At each step, the variable that minimizes the overall Wilks' Lambda is entered.

- Maximum number of steps is 18.
- Minimum partial F to enter is 3.84.
- Maximum partial F to remove is 2.71.
- F level, tolerance, or VIN insufficient for further computation.

Variables Entered/Removed^{a,b,c,d}

Step	Wilks' Lambda			
	Exact F			
	Statistic	df1	df2	Sig.
1	119.257	2	95.000	.000

2	71.420	4	188.000	.000
3	55.553	6	186.000	.000
4	49.257	8	184.000	.000
5	43.576	10	182.000	.000
6	38.877	12	180.000	.000
7	35.116	14	178.000	.000

At each step, the variable that minimizes the overall Wilks' Lambda is entered.

- Maximum number of steps is 18.
- Minimum partial F to enter is 3.84.
- Maximum partial F to remove is 2.71.
- F level, tolerance, or VIN insufficient for further computation.

Variables in the Analysis

Step		Tolerance	F to Remove	Wilks' Lambda
1	IDH-2014_n	1.000	119.257	
2	IDH-2014_n	.932	127.404	.585
	Religious tolerance_n	.932	37.989	.285
3	IDH-2014_n	.914	105.993	.421
	Religious tolerance_n	.904	37.974	.233
	Democracy Index-2014_n	.935	10.601	.158
4	IDH-2014_n	.785	92.030	.304
	Religious tolerance_n	.712	55.619	.224
	Democracy Index-2014_n	.471	25.133	.157
	Governance	.429	12.240	.128
5	IDH-2014_n	.570	11.295	.108
	Religious tolerance_n	.712	52.212	.186
	Democracy Index-2014_n	.465	22.051	.129
	Governance	.428	11.500	.109
	EPI-2014_n	.650	7.614	.101
6	IDH-2014_n	.556	10.758	.096
	Religious tolerance_n	.709	39.398	.145
	Democracy Index-2014_n	.441	25.218	.121
	Governance	.406	13.882	.101
	EPI-2014_n	.649	7.542	.091
	Tolerance for immigrants_n	.887	5.388	.087
7	IDH-2014_n	.523	6.599	.081
	Religious tolerance_n	.709	38.959	.133
	Democracy Index-2014_n	.439	23.554	.108

Governance	.406	13.678	.092
EPI-2014_n	.616	9.254	.085
Tolerance for immigrants_n	.872	4.387	.078
Water withdrawals_np	.893	4.317	.078

Variables Not in the Analysis

Step		Tolerance	Min. Tolerance	F to Enter	Wilks' Lambda
0	Opportunity	1.000	1.000	56.981	.455
	Water withdrawals_np	1.000	1.000	23.286	.671
	Tolerance for immigrants_n	1.000	1.000	14.241	.769
	Religious tolerance_n	1.000	1.000	33.762	.585
	IDH-2014_n	1.000	1.000	119.257	.285
	Governance	1.000	1.000	13.585	.778
	EPI-2014_n	1.000	1.000	116.304	.290
	HPI	1.000	1.000	13.512	.779
Democracy Index-2014_n	1.000	1.000	16.196	.746	
1	Opportunity	.931	.931	25.626	.184
	Water withdrawals_np	.976	.976	6.027	.252
	Tolerance for immigrants_n	.979	.979	9.951	.235
	Religious tolerance_n	.932	.932	37.989	.158
	Governance	.903	.903	.718	.281
	EPI-2014_n	.665	.665	11.645	.228
	HPI	.980	.980	3.094	.267
	Democracy Index-2014_n	.965	.965	10.449	.233
2	Opportunity	.867	.834	7.160	.137
	Water withdrawals_np	.976	.910	5.019	.142
	Tolerance for immigrants_n	.943	.898	2.924	.148
	Governance	.850	.809	.251	.157
	EPI-2014_n	.662	.643	9.823	.130
	HPI	.952	.906	4.478	.144
	Democracy Index-2014_n	.935	.904	10.601	.128
	3	Opportunity	.431	.431	1.933
Water withdrawals_np		.959	.887	3.689	.119
Tolerance for immigrants_n		.935	.876	3.299	.120
Governance		.429	.429	12.240	.101
EPI-2014_n		.650	.641	8.269	.109
HPI		.944	.883	2.953	.121
4	Opportunity	.384	.374	3.961	.093
	Water withdrawals_np	.958	.428	3.665	.094

	Tolerance for immigrants_n	.888	.407	5.437	.091
	EPI-2014_n	.650	.428	7.614	.087
	HPI	.871	.396	1.005	.099
5	Opportunity	.382	.372	2.759	.082
	Water withdrawals_np	.908	.428	5.316	.078
	Tolerance for immigrants_n	.887	.406	5.388	.078
	HPI	.866	.395	1.261	.084
6	Opportunity	.310	.310	2.986	.073
	Water withdrawals_np	.893	.406	4.317	.071
	HPI	.851	.381	1.378	.075
7	Opportunity	.310	.310	2.895	.066
	HPI	.847	.381	1.165	.069

Wilks' Lambda

Step	Number of Variables	Lambda	df1	df2	df3
1	1	.285	1	2	95
2	2	.158	2	2	95
3	3	.128	3	2	95
4	4	.101	4	2	95
5	5	.087	5	2	95
6	6	.078	6	2	95
7	7	.071	7	2	95

Wilks' Lambda

Step	Exact F			
	Statistic	df1	df2	Sig.
1	119.257	2	95.000	.000
2	71.420	4	188.000	.000
3	55.553	6	186.000	.000
4	49.257	8	184.000	.000
5	43.576	10	182.000	.000
6	38.877	12	180.000	.000
7	35.116	14	178.000	.000

Summary of Canonical Discriminant Functions

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation

1	3.945 ^a	67.9	67.9	.893
2	1.862 ^a	32.1	100.0	.807

a. First 2 canonical discriminant functions were used in the analysis.

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	.071	243.789	14	.000
2	.349	96.739	6	.000

Standardized Canonical Discriminant Function

Coefficients

	Function	
	1	2
Water withdrawals_np	-.274	.246
Tolerance for immigrants_n	.065	.391
Religious tolerance_n	.512	.831
IDH-2014_n	.552	.076
Governance	-.347	-.861
EPI-2014_n	.582	-.120
Democracy Index-2014_n	.385	1.015

Structure Matrix

	Function	
	1	2
EPI-2014_n	.774 [*]	-.216
IDH-2014_n	.753 [*]	-.382
Opportunity ^a	.369 [*]	.303
Governance	.267 [*]	-.045
Democracy Index-2014_n	.246 [*]	.234
HPI ^a	.110 [*]	-.036
Religious tolerance_n	.129	.589 [*]
Tolerance for immigrants_n	-.071	.388 [*]
Water withdrawals_np	-.257	.351 [*]

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions

Variables ordered by absolute size of correlation within function.

*. Largest absolute correlation between each variable and any discriminant function

a. This variable not used in the analysis.

Functions at Group Centroids

ClusterM_reOrd	Function	
	1	2
1	-2.494	.417
2	2.088	1.567
3	.921	-1.731

Unstandardized canonical discriminant functions evaluated at group means

Classification Statistics

Classification Processing Summary

Processed	132
Excluded	0
Missing or out-of-range group codes	
At least one missing discriminating variable	0
Used in Output	132

Prior Probabilities for Groups

ClusterM_reOrd	Prior	Cases Used in Analysis	
		Unweighted	Weighted
1	.333	36	36.000
2	.333	28	28.000
3	.333	34	34.000
Total	1.000	98	98.000

Classification Results^a

ClusterM_reOrd			Predicted Group Membership			Total
			1	2	3	
Original Count	1	33	1	2	36	
	2	0	28	0	28	

	3	2	0	32	34
	Ungrouped cases	0	20	14	34
%	1	91.7	2.8	5.6	100.0
	2	.0	100.0	.0	100.0
	3	5.9	.0	94.1	100.0
	Ungrouped cases	.0	58.8	41.2	100.0

a. 94.9% of original grouped cases correctly classified.

RESULTADOS: Análise Discriminante: by GROUPS=3Reg_reOrd

DISCRIMINANT

```

/GROUPS=@3Reg_reOrd(1 3)
/VARIABLES=Opportunity_n Waterwithdrawals_np Toleranceforimmigrants_n Religioustolerance_n
IDH2014_n Governança_n EPI2014_n HPI_n DemocracyIndex2014_n
/ANALYSIS ALL
/METHOD=WILKS
/FIN=3.84
/FOOT=2.71
/PRIORS EQUAL
/HISTORY
/STATISTICS=TABLE
/CLASSIFY=NONMISSING POOLED.

```

Discriminant

		Notes
Output Created		22-May-2016 15:35:59
Comments		
Input	Data	C:\Users\ Arvores de Classificacao.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data	132
	File	
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing in the analysis phase.
	Cases Used	In the analysis phase, cases with no user- or system-missing values for any predictor variable are used. Cases with user-, system-missing, or out-of-range values for the grouping variable are always excluded.

Syntax	DISCRIMINANT /GROUPS=@3Reg_reOrd(1 3) /VARIABLES=Opportunity_n Waterwithdrawals_np Toleranceforimmigrants_n Religioustolerance_n IDH2014_n Governança_n EPI2014_n HPI_n DemocracyIndex2014_n /ANALYSIS ALL /METHOD=WILKS /FIN=3.84 /FOUT=2.71 /PRIORS EQUAL /HISTORY /STATISTICS=TABLE /CLASSIFY=NONMISSING POOLED.	
Resources	Processor Time	00:00:00.046
	Elapsed Time	00:00:00.128

[DataSet1] C:\Users\Arvores de Classificacao.sav

Analysis Case Processing Summary

Unweighted Cases		N	Percent
Valid		132	100.0
Excluded	Missing or out-of-range group codes	0	.0
	At least one missing discriminating variable	0	.0
	Both missing or out-of-range group codes and at least one missing discriminating variable	0	.0
	Total	0	.0
Total		132	100.0

Group Statistics

3Reg_reOrd		Valid N (listwise)	
		Unweighted	Weighted
1	Opportunity	84	84.000
	Water withdrawals_np	84	84.000
	Tolerance for immigrants_n	84	84.000

	Religious tolerance_n	84	84.000
	IDH-2014_n	84	84.000
	Governance	84	84.000
	EPI-2014_n	84	84.000
	HPI	84	84.000
	Democracy Index-2014_n	84	84.000
2	Opportunity	21	21.000
	Water withdrawals_np	21	21.000
	Tolerance for immigrants_n	21	21.000
	Religious tolerance_n	21	21.000
	IDH-2014_n	21	21.000
	Governance	21	21.000
	EPI-2014_n	21	21.000
	HPI	21	21.000
	Democracy Index-2014_n	21	21.000
3	Opportunity	27	27.000
	Water withdrawals_np	27	27.000
	Tolerance for immigrants_n	27	27.000
	Religious tolerance_n	27	27.000
	IDH-2014_n	27	27.000
	Governance	27	27.000
	EPI-2014_n	27	27.000
	HPI	27	27.000
	Democracy Index-2014_n	27	27.000
Total	Opportunity	132	132.000
	Water withdrawals_np	132	132.000
	Tolerance for immigrants_n	132	132.000
	Religious tolerance_n	132	132.000
	IDH-2014_n	132	132.000
	Governance	132	132.000
	EPI-2014_n	132	132.000
	HPI	132	132.000
	Democracy Index-2014_n	132	132.000

Analysis 1

Stepwise Statistics

Variables Entered/Removed^{a,b,c,d}

Step	Entered	Wilks' Lambda			
		Statistic	df1	df2	df3
1	Opportunity	.382	1	2	129.000
2	Governance	.274	2	2	129.000
3	HPI	.222	3	2	129.000
4	Religious tolerance_n	.199	4	2	129.000

At each step, the variable that minimizes the overall Wilks' Lambda is entered.

- Maximum number of steps is 18.
- Minimum partial F to enter is 3.84.
- Maximum partial F to remove is 2.71.
- F level, tolerance, or VIN insufficient for further computation.

Variables Entered/Removed^{a,b,c,d}

Step	Wilks' Lambda			
	Exact F			
	Statistic	df1	df2	Sig.
1	104.392	2	129.000	.000
2	58.351	4	256.000	.000
3	47.451	6	254.000	.000
4	39.140	8	252.000	.000

At each step, the variable that minimizes the overall Wilks' Lambda is entered.

- Maximum number of steps is 18.
- Minimum partial F to enter is 3.84.
- Maximum partial F to remove is 2.71.
- F level, tolerance, or VIN insufficient for further computation.

Variables in the Analysis

Step		Tolerance	F to Remove	Wilks' Lambda
1	Opportunity	1.000	104.392	
2	Opportunity	.288	29.405	.399
	Governance	.288	25.327	.382
3	Opportunity	.288	21.490	.298
	Governance	.288	20.409	.294
	HPI	.999	14.655	.274
4	Opportunity	.272	12.934	.240
	Governance	.285	14.460	.244
	HPI	.905	19.357	.260

Variables in the Analysis

Step		Tolerance	F to Remove	Wilks' Lambda
1	Opportunity	1.000	104.392	
2	Opportunity	.288	29.405	.399
	Governance	.288	25.327	.382
3	Opportunity	.288	21.490	.298
	Governance	.288	20.409	.294
	HPI	.999	14.655	.274
4	Opportunity	.272	12.934	.240
	Governance	.285	14.460	.244
	HPI	.905	19.357	.260
	Religious tolerance_n	.832	7.434	.222

Variables Not in the Analysis

Step		Tolerance	Min. Tolerance	F to Enter	Wilks' Lambda
0	Opportunity	1.000	1.000	104.392	.382
	Water withdrawals_np	1.000	1.000	.060	.999
	Tolerance for immigrants_n	1.000	1.000	9.739	.869
	Religious tolerance_n	1.000	1.000	9.818	.868
	IDH-2014_n	1.000	1.000	51.614	.555
	Governance	1.000	1.000	97.018	.399
	EPI-2014_n	1.000	1.000	40.281	.616
	HPI	1.000	1.000	22.688	.740
	Democracy Index-2014_n	1.000	1.000	61.189	.513
1	Water withdrawals_np	.992	.992	.519	.379
	Tolerance for immigrants_n	.914	.914	1.332	.374
	Religious tolerance_n	.927	.927	6.975	.344
	IDH-2014_n	.675	.675	1.355	.374
	Governance	.288	.288	25.327	.274
	EPI-2014_n	.674	.674	.285	.380
	HPI	1.000	1.000	19.201	.294
	Democracy Index-2014_n	.475	.475	.152	.381
2	Water withdrawals_np	.988	.286	.570	.271
	Tolerance for immigrants_n	.898	.267	.188	.273
	Religious tolerance_n	.919	.274	3.340	.260
	IDH-2014_n	.654	.274	.784	.270
	EPI-2014_n	.668	.265	.595	.271
	HPI	.999	.288	14.655	.222
	Democracy Index-2014_n	.432	.261	1.497	.267

3	Water withdrawals_np	.900	.285	2.973	.212
	Tolerance for immigrants_n	.795	.264	2.694	.213
	Religious tolerance_n	.832	.272	7.434	.199
	IDH-2014_n	.580	.273	.868	.219
	EPI-2014_n	.630	.264	.020	.222
	Democracy Index-2014_n	.421	.260	.449	.221
4	Water withdrawals_np	.880	.267	2.024	.193
	Tolerance for immigrants_n	.791	.252	1.939	.193
	IDH-2014_n	.524	.246	.111	.198
	EPI-2014_n	.599	.241	.248	.198
	Democracy Index-2014_n	.412	.241	1.012	.196

Wilks' Lambda

Step	Number of Variables	Lambda	df1	df2	df3
1	1	.382	1	2	129
2	2	.274	2	2	129
3	3	.222	3	2	129
4	4	.199	4	2	129

Wilks' Lambda

Step	Exact F			
	Statistic	df1	df2	Sig.
1	104.392	2	129.000	.000
2	58.351	4	256.000	.000
3	47.451	6	254.000	.000
4	39.140	8	252.000	.000

Summary of Canonical Discriminant Functions

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	1.770 ^a	68.5	68.5	.799
2	.815 ^a	31.5	100.0	.670

a. First 2 canonical discriminant functions were used in the analysis.

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	.199	205.939	8	.000
2	.551	76.013	3	.000

Standardized Canonical Discriminant Function**Coefficients**

	Function	
	1	2
Opportunity	.737	.789
Religious tolerance_n	.025	.531
Governance	.256	-1.169
HPI	.273	.687

Structure Matrix

	Function	
	1	2
Opportunity	.955*	-.066
Governance	.875*	-.427
Democracy Index-2014_n ^a	.744*	-.182
IDH-2014_n ^a	.630*	-.106
EPI-2014_n ^a	.603*	-.070
Tolerance for immigrants_n ^a	.179*	-.068
Water withdrawals_np ^a	-.156*	-.124
HPI	.246	.548*
Religious tolerance_n	.188	.332*

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions

Variables ordered by absolute size of correlation within function.

*. Largest absolute correlation between each variable and any discriminant function

a. This variable not used in the analysis.

Functions at Group Centroids

3Reg_reOrd	Function	
	1	2
1	-.934	-.231
2	.789	1.981

3	2.293	-.823
---	-------	-------

Unstandardized canonical discriminant functions evaluated at group means

Classification Statistics

Classification Processing Summary

Processed	132
Excluded	0
Missing or out-of-range group codes	
At least one missing discriminating variable	0
Used in Output	132

Prior Probabilities for Groups

3Reg_reOrd	Prior	Cases Used in Analysis	
		Unweighted	Weighted
1	.333	84	84.000
2	.333	21	21.000
3	.333	27	27.000
Total	1.000	132	132.000

Classification Results^a

3Reg_reOrd		Predicted Group Membership			Total
		1	2	3	
Original	Count	1	2	3	
	1	74	6	4	84
	2	0	17	4	21
	3	2	0	25	27
%	1	88.1	7.1	4.8	100.0
	2	.0	81.0	19.0	100.0
	3	7.4	.0	92.6	100.0

a. 87.9% of original grouped cases correctly classified.

11.3 Resultados das análises discriminantes (Minitab e SPSS)

Nesta sessão, apresentamos uma tabela contendo o conjunto dos resultados dos testes de modelos realizados na sessão 10 e os resultados obtidos neste trabalho, onde foi utilizado o *software* SPSS, para fins de comparação.

Modelo	ClusterM_reOrd	3Reg_reOrd
AD (Análise Discriminante) Minitab	99.2% (Q)	95.5% (Q)
RL (Regressão Logística) Minitab	94.9% (O)	95.2% (O)
AD (Análise Discriminante) SPSS	94.9%	87.9%

legenda:

ClusterM_reOrd = Agrupamento de 4 clusters ordenados, sendo de 1 (pior) a 4 (melhor)

3Reg_reOrd = As 3 regiões ordenadas, sendo 1 para Others, 2 AIBER e 3 para AVECO

Q = método *quadratic* na análise discriminante

O = método *ordinal* na regressão logística

Minitab = resultados obtidos através do *software* Minitab

SPSS = resultados obtidos através do *software* SPSS

Na tabela acima vemos que o modelo de análise discriminante realizado no *software* Minitab, apresentou melhores resultados que o realizado no SPSS.

11.4 Árvores de Classificação

Nesta sessão, apresentaremos análises de árvores de classificação, ou árvores de decisão.

A árvore de decisão consiste de uma hierarquia de nós internos e externos que são conectados por ramos. O nó interno, também conhecido como nó decisório ou nó intermediário, é a unidade de tomada de decisão que avalia através de teste lógico qual será o próximo nó descendente ou filho. Em contraste, um nó externo (não tem nó descendente), também conhecido como folha ou nó terminal, está associado a um rótulo ou a um valor.

Em geral, o procedimento de uma árvore de decisão é o seguinte: apresenta-se um conjunto de dados ao nó inicial (ou nó raiz que também é um nó interno) da árvore; dependendo do resultado do teste lógico usado pelo nó, a árvore ramifica-se para um dos nós filhos e este procedimento é repetido até que um nó terminal é alcançado. A repetição deste procedimento caracteriza a recursividade da árvore de decisão. *fonte: Sistema Maxwell – PUC-Rio.*
http://www.maxwell.vrac.puc-rio.br/3710/3710_4.PDF

11.4.1 Árvores de Classificação – utilizando ClusterM_reOrd

Nesta sessão, apresentaremos análises de árvores de classificação, ou árvores de decisão, tendo ClusterM_reOrd como variável dependente.

Os métodos explorados serão CHAID, Exhaustive CHAID e CRT, conforme abaixo:

Método: CHAID

* Decision Tree.


```

TREE ClusterM_reOrd [s] BY Opportunity_n [s] Waterwithdrawals_np [s] Toleranceforimmigrants_n [s]
Religioustolerance_n [s] IDH2014_n [s] Governança_n [s] EPI2014_n [s] HPI_n [s] DemocracyIndex2014_n [s]
/TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES NODEDEFS=YES
SCALE=AUTO
/PRINT MODELSUMMARY RISK
/GAIN SUMMARYTABLE=YES TYPE=[NODE] SORT=DESCENDING CUMULATIVE=NO
/METHOD TYPE=CHAID
/GROWTHLIMIT MAXDEPTH=5 MINPARENTSIZE=10 MINCHILDSIZE=5
/VALIDATION TYPE=NONE OUTPUT=BOTHSAMPLES
/CHAID ALPHASPLIT=0.05 ALPHAMERGE=0.05 SPLITMERGED=NO ADJUST=BONFERRONI
INTERVALS=10.

```

Classification Tree

Notes		
Output Created		22-May-2016 16:30:25
Comments		
Input	Data	C:\Users\ Arvores de Classificacao.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data	132
	File	
Missing Value Handling	Definition of Missing	Handling of user-defined missing values of nominal independent variables depends on the growing method.
	Cases Used	Only cases with valid data for the dependent variable and some or all independent variables are used in computing any statistics.

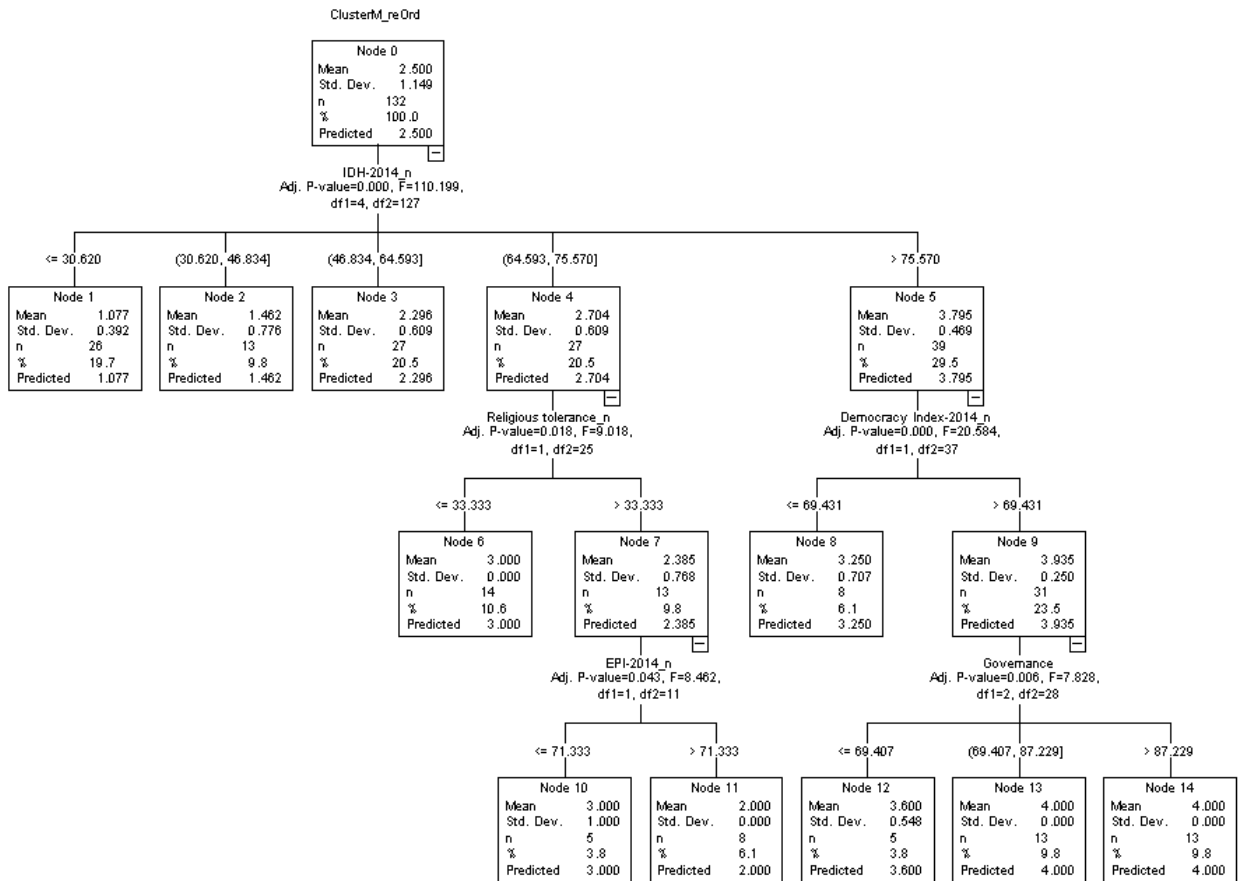
Syntax	<pre> TREE ClusterM_reOrd [s] BY Opportunity_n [s] Waterwithdrawals_np [s] Toleranceforimmigrants_n [s] Religioustolerance_n [s] IDH2014_n [s] Governança_n [s] EPI2014_n [s] HPI_n [s] DemocracyIndex2014_n [s] /TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES NODEDEFS=YES SCALE=AUTO /PRINT MODELSUMMARY RISK /GAIN SUMMARYTABLE=YES TYPE=[NODE] SORT=DESCENDING CUMULATIVE=NO /METHOD TYPE=CHAID /GROWTHLIMIT MAXDEPTH=5 MINPARENTSIZE=10 MINCHILDSIZE=5 /VALIDATION TYPE=NONE OUTPUT=BOTHSAMPLES /CHAID ALPHASPLIT=0.05 ALPHAMERGE=0.05 SPLITMERGED=NO ADJUST=BONFERRONI INTERVALS=10. </pre>	
Resources	Processor Time	00:00:00.515
Files Saved	Rules File	

[DataSet1] C:\Users\Arvores de Classificacao.sav

Model Summary

Specifications	Growing Method	CHAID
	Dependent Variable	ClusterM_reOrd
	Independent Variables	Opportunity, Water withdrawals_np, Tolerance for immigrants_n, Religious tolerance_n, IDH-2014_n, Governance, EPI-2014_n, HPI, Democracy Index-2014_n
	Validation	None
	Maximum Tree Depth	5
	Minimum Cases in Parent Node	10
	Minimum Cases in Child Node	5

Results	Independent Variables	IDH-2014_n, Religious tolerance_n, EPI-2014_n,
	Included	Democracy Index-2014_n, Governance
	Number of Nodes	15
	Number of Terminal Nodes	10
	Depth	3



Gain Summary for Nodes

Node	N	Percent	Mean
13	13	9.8%	4.00
14	13	9.8%	4.00
12	5	3.8%	3.60
8	8	6.1%	3.25
6	14	10.6%	3.00
10	5	3.8%	3.00
3	27	20.5%	2.30
11	8	6.1%	2.00
2	13	9.8%	1.46
1	26	19.7%	1.08

Gain Summary for Nodes

Node	N	Percent	Mean
13	13	9.8%	4.00
14	13	9.8%	4.00
12	5	3.8%	3.60
8	8	6.1%	3.25
6	14	10.6%	3.00
10	5	3.8%	3.00
3	27	20.5%	2.30
11	8	6.1%	2.00
2	13	9.8%	1.46
1	26	19.7%	1.08

Growing Method: CHAID

Dependent Variable: ClusterM_reOrd

Risk

Estimate	Std. Error
.223	.045

Growing Method: CHAID

Dependent Variable:

ClusterM_reOrd

Método: Exhaustive CHAID

* Decision Tree.

```

TREE ClusterM_reOrd [s] BY Opportunity_n [s] Waterwithdrawals_np [s] Toleranceforimmigrants_n [s]
Religioustolerance_n [s] IDH2014_n [s] Governança_n [s] EPI2014_n [s] HPI_n [s] DemocracyIndex2014_n [s]
/TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES NODEDEFS=YES
SCALE=AUTO
/PRINT MODELSUMMARY RISK
/GAIN SUMMARYTABLE=YES TYPE=[NODE] SORT=DESCENDING CUMULATIVE=NO
/METHOD TYPE=EXHAUSTIVECHAID
/GROWTHLIMIT MAXDEPTH=5 MINPARENTSIZE=10 MINCHILDSIZE=5
/VALIDATION TYPE=NONE OUTPUT=BOTHSAMPLES
/CHAID ALPHASPLIT=0.05 SPLITMERGED=NO ADJUST=BONFERRONI INTERVALS=10.

```

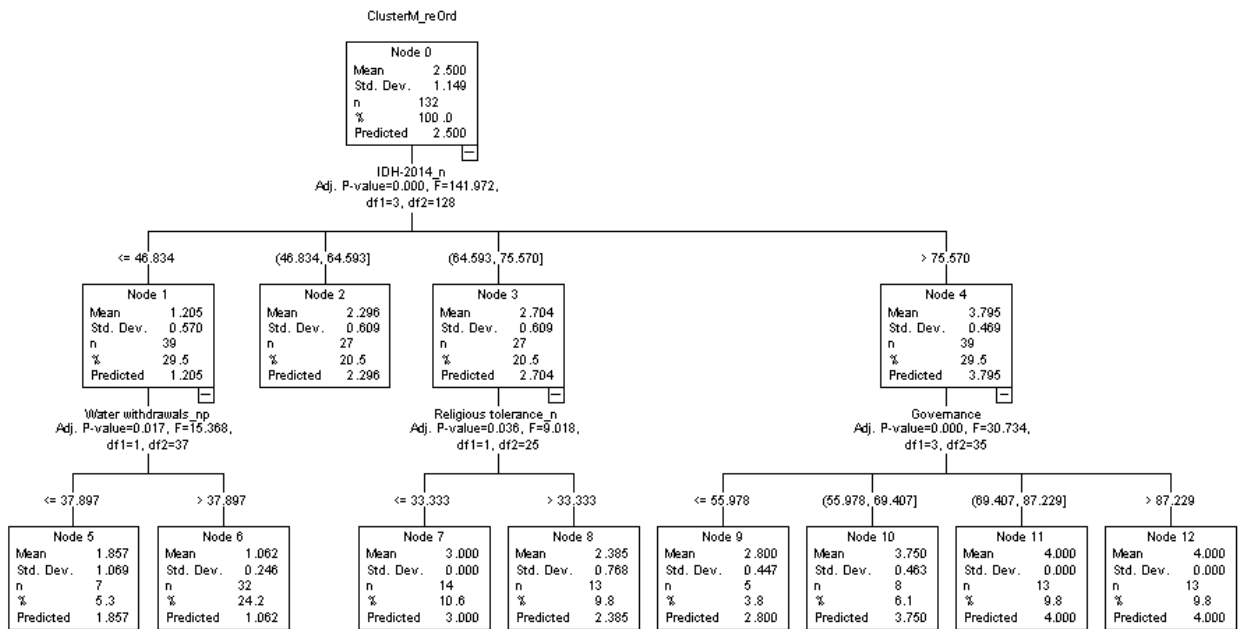
Classification Tree**Notes**

Output Created		22-May-2016 16:30:57
Comments		
Input	Data	C:\Users\Arvores de Classificacao.sav
	Active Dataset	DataSet1

	Filter	<none>	
	Weight	<none>	
	Split File	<none>	
	N of Rows in Working Data		132
Missing Value Handling	File		
	Definition of Missing	Handling of user-defined missing values of nominal independent variables depends on the growing method.	
	Cases Used	Only cases with valid data for the dependent variable and some or all independent variables are used in computing any statistics.	
Syntax		<pre> TREE ClusterM_reOrd [s] BY Opportunity_n [s] Waterwithdrawals_np [s] Toleranceforimmigrants_n [s] Religioustolerance_n [s] IDH2014_n [s] Governança_n [s] EPI2014_n [s] HPI_n [s] DemocracyIndex2014_n [s] /TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES NODEDEFS=YES SCALE=AUTO /PRINT MODELSUMMARY RISK /GAIN SUMMARYTABLE=YES TYPE=[NODE] SORT=DESCENDING CUMULATIVE=NO /METHOD TYPE=EXHAUSTIVECHAID /GROWTHLIMIT MAXDEPTH=5 MINPARENTSIZE=10 MINCHILDSIZE=5 /VALIDATION TYPE=NONE OUTPUT=BOTHSAMPLES /CHAID ALPHASPLIT=0.05 SPLITMERGED=NO ADJUST=BONFERRONI INTERVALS=10. </pre>	
Resources	Processor Time		00:00:00.717
	Elapsed Time		00:00:01.986
Files Saved	Rules File		

Model Summary

Specifications	Growing Method	EXHAUSTIVE CHAID
	Dependent Variable	ClusterM_reOrd
	Independent Variables	Opportunity, Water withdrawals_np, Tolerance for immigrants_n, Religious tolerance_n, IDH-2014_n, Governance, EPI-2014_n, HPI, Democracy Index-2014_n
	Validation	None
	Maximum Tree Depth	5
	Minimum Cases in Parent Node	10
	Minimum Cases in Child Node	5
Results	Independent Variables Included	IDH-2014_n, Water withdrawals_np, Religious tolerance_n, Governance
	Number of Nodes	13
	Number of Terminal Nodes	9
	Depth	2



Gain Summary for Nodes

Node	N	Percent	Mean
11	13	9.8%	4.00
12	13	9.8%	4.00
10	8	6.1%	3.75

7	14	10.6%	3.00
9	5	3.8%	2.80
8	13	9.8%	2.38
2	27	20.5%	2.30
5	7	5.3%	1.86
6	32	24.2%	1.06

Growing Method: EXHAUSTIVE CHAID

Dependent Variable: ClusterM_reOrd

Risk

Estimate	Std. Error
.210	.039

Growing Method:

EXHAUSTIVE CHAID

Dependent Variable:

ClusterM_reOrd

Método: CRT

* Decision Tree.

```

TREE ClusterM_reOrd [s] BY Opportunity_n [s] Waterwithdrawals_np [s] Toleranceforimmigrants_n [s]
Religioustolerance_n [s] IDH2014_n [s] Governança_n [s] EPI2014_n [s] HPI_n [s] DemocracyIndex2014_n [s]
/TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES NODEDEFS=YES
SCALE=AUTO
/PRINT MODELSUMMARY RISK
/GAIN SUMMARYTABLE=YES TYPE=[NODE] SORT=DESCENDING CUMULATIVE=NO
/METHOD TYPE=CRT MAXSURROGATES=AUTO PRUNE=NONE
/GROWTHLIMIT MAXDEPTH=5 MINPARENTSIZE=10 MINCHILDSIZE=5
/VALIDATION TYPE=NONE OUTPUT=BOTHSAMPLES
/CRT MINIMPROVEMENT=0.0001.
    
```

Classification Tree

Notes

Output Created	22-May-2016 16:31:31
Comments	
Input	Data
	Active Dataset
	Filter
	Weight
	Split File
	N of Rows in Working Data
	File
	132

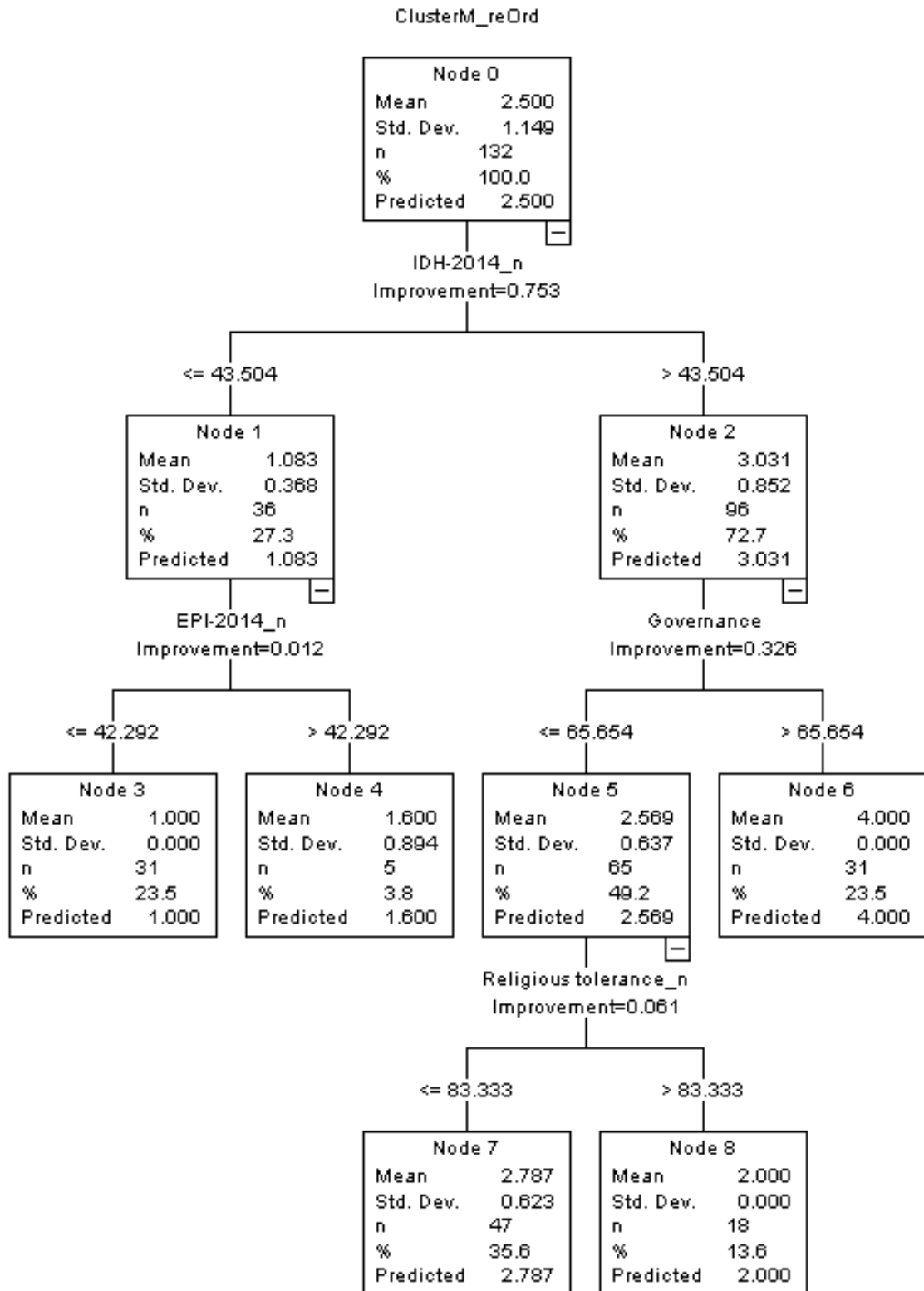
Missing Value Handling	Definition of Missing	Handling of user-defined missing values of nominal independent variables depends on the growing method.
	Cases Used	Only cases with valid data for the dependent variable and some or all independent variables are used in computing any statistics.
Syntax		<pre> TREE ClusterM_reOrd [s] BY Opportunity_n [s] Waterwithdrawals_np [s] Toleranceforimmigrants_n [s] Religioustolerance_n [s] IDH2014_n [s] Governana_n [s] EPI2014_n [s] HPI_n [s] DemocracyIndex2014_n [s] /TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES NODEDEFS=YES SCALE=AUTO /PRINT MODELSUMMARY RISK /GAIN SUMMARYTABLE=YES TYPE=[NODE] SORT=DESCENDING CUMULATIVE=NO /METHOD TYPE=CRT MAXSURROGATES=AUTO PRUNE=NONE /GROWTHLIMIT MAXDEPTH=5 MINPARENTSIZE=10 MINCHILDSIZE=5 /VALIDATION TYPE=NONE OUTPUT=BOTHSAMPLES /CRT MINIMPROVEMENT=0.0001. </pre>
Resources	Processor Time	00:00:00.531
	Elapsed Time	00:00:01.763
Files Saved	Rules File	

[DataSet1] C:\Users\Arvores de Classificacao.sav

Model Summary

Specifications	Growing Method	CRT
	Dependent Variable	ClusterM_reOrd

	Independent Variables	Opportunity, Water withdrawals_np, Tolerance for immigrants_n, Religious tolerance_n, IDH-2014_n, Governance, EPI-2014_n, HPI, Democracy Index-2014_n	
	Validation	None	
	Maximum Tree Depth		5
	Minimum Cases in Parent Node		10
	Minimum Cases in Child Node		5
Results	Independent Variables Included	IDH-2014_n, EPI-2014_n, Opportunity, Water withdrawals_np, Governance, HPI, Democracy Index-2014_n, Tolerance for immigrants_n, Religious tolerance_n	
	Number of Nodes		9
	Number of Terminal Nodes		5
	Depth		3



Gain Summary for Nodes

Node	N	Percent	Mean
6	31	23.5%	4.00
7	47	35.6%	2.79
8	18	13.6%	2.00
4	5	3.8%	1.60
3	31	23.5%	1.00

Growing Method: CRT

Dependent Variable: ClusterM_reOrd

Risk

Estimate	Std. Error
.160	.042

Growing Method: CRT

Dependent Variable:

ClusterM_reOrd

11.4.2 Árvores de Classificação – utilizando “3Reg_reOrd”

Nesta sessão, apresentaremos análises de árvores de classificação, ou árvores de decisão, tendo 3Reg_reOrd como variável dependente. Os métodos explorados serão CHAID, Exhaustive CHAID e CRT, conforme abaixo:

Método: CHAID

* Decision Tree.

```

TREE @3Reg_reOrd [s] BY Opportunity_n [s] Waterwithdrawals_np [s] Toleranceforimmigrants_n [s]
Religioustolerance_n [s] IDH2014_n [s] Governança_n [s] EPI2014_n [s] HPI_n [s] DemocracyIndex2014_n [s]
/TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES NODEDEFS=YES
SCALE=AUTO
/PRINT MODELSUMMARY RISK
/GAIN SUMMARYTABLE=YES TYPE=[NODE] SORT=DESCENDING CUMULATIVE=NO
/METHOD TYPE=CHAID
/GROWTHLIMIT MAXDEPTH=5 MINPARENTSIZE=10 MINCHILDSIZE=5
/VALIDATION TYPE=NONE OUTPUT=BOTHSAMPLES
/CHAID ALPHASPLIT=0.05 ALPHAMERGE=0.05 SPLITMERGED=NO ADJUST=BONFERRONI
INTERVALS=10.

```

Classification Tree

Notes

Output Created		22-May-2016 16:32:42
Comments		
Input	Data	C:\Users\Cla\Arvores de Classificacao.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	132
Missing Value Handling	Definition of Missing	Handling of user-defined missing values of nominal independent variables depends on the growing method.
	Cases Used	Only cases with valid data for the dependent variable and some or all independent variables are used in computing any statistics.

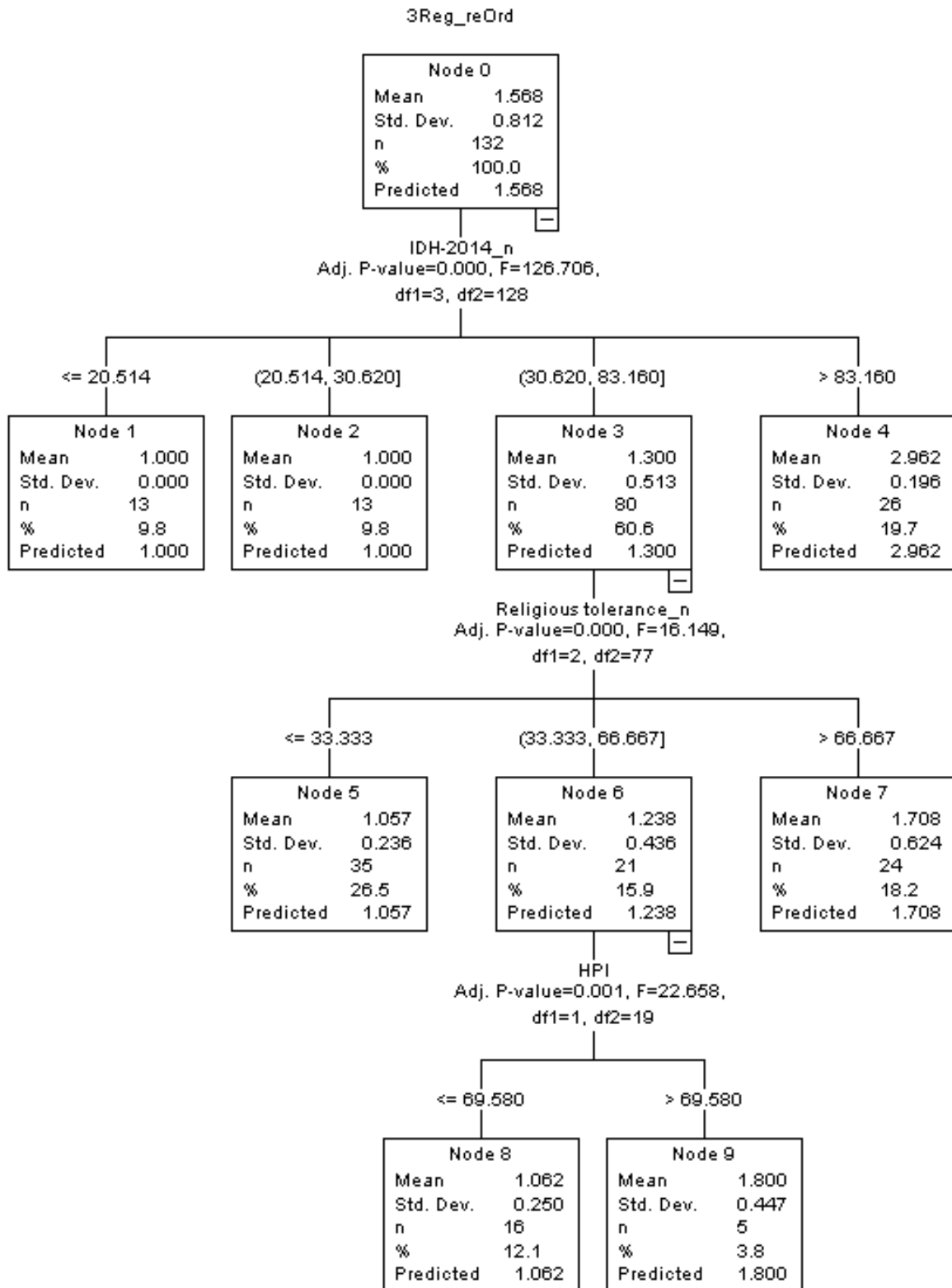
Syntax	<pre> TREE @3Reg_reOrd [s] BY Opportunity_n [s] Waterwithdrawals_np [s] Toleranceforimmigrants_n [s] Religioustolerance_n [s] IDH2014_n [s] Governança_n [s] EPI2014_n [s] HPI_n [s] DemocracyIndex2014_n [s] /TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES NODEDEFS=YES SCALE=AUTO /PRINT MODELSUMMARY RISK /GAIN SUMMARYTABLE=YES TYPE=[NODE] SORT=DESCENDING CUMULATIVE=NO /METHOD TYPE=CHAID /GROWTHLIMIT MAXDEPTH=5 MINPARENTSIZE=10 MINCHILDSIZE=5 /VALIDATION TYPE=NONE OUTPUT=BOTHSAMPLES /CHAID ALPHASPLIT=0.05 ALPHAMERGE=0.05 SPLITMERGED=NO ADJUST=BONFERRONI INTERVALS=10. </pre>	
Resources	Processor Time	00:00:00.593
	Elapsed Time	00:00:01.580
Files Saved	Rules File	

[DataSet1] C:\Users\Arvores de Classificacao.sav

Model Summary

Specifications	Growing Method	CHAID
	Dependent Variable	3Reg_reOrd
	Independent Variables	Opportunity, Water withdrawals_np, Tolerance for immigrants_n, Religious tolerance_n, IDH-2014_n, Governance, EPI-2014_n, HPI, Democracy Index-2014_n
	Validation	None
	Maximum Tree Depth	5
	Minimum Cases in Parent Node	10

Results	Minimum Cases in Child Node		5
	Independent Variables Included	IDH-2014_n, Religious tolerance_n, HPI	
	Number of Nodes		10
	Number of Terminal Nodes		7
	Depth		3



Gain Summary for Nodes

Node	N	Percent	Mean
4	26	19.7%	2.96
9	5	3.8%	1.80
7	24	18.2%	1.71
8	16	12.1%	1.06
5	35	26.5%	1.06
1	13	9.8%	1.00
2	13	9.8%	1.00

Growing Method: CHAID

Dependent Variable: 3Reg_reOrd

Risk

Estimate	Std. Error
.103	.024

Growing Method: CHAID

Dependent Variable:

3Reg_reOrd

Método: Exhaustive CHAID

* Decision Tree.

```

TREE @3Reg_reOrd [s] BY Opportunity_n [s] Waterwithdrawals_np [s] Toleranceforimmigrants_n [s]
Religioustolerance_n [s] IDH2014_n [s] Governança_n [s] EPI2014_n [s] HPI_n [s] DemocracyIndex2014_n [s]
/TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES NODEDEFS=YES
SCALE=AUTO
/PRINT MODELSUMMARY RISK
/GAIN SUMMARYTABLE=YES TYPE=[NODE] SORT=DESCENDING CUMULATIVE=NO
/METHOD TYPE=EXHAUSTIVECHAID
/GROWTHLIMIT MAXDEPTH=5 MINPARENTSIZE=10 MINCHILDSIZE=5
/VALIDATION TYPE=NONE OUTPUT=BOTHSAMPLES
/CHAID ALPHASPLIT=0.05 SPLITMERGED=NO ADJUST=BONFERRONI INTERVALS=10.

```

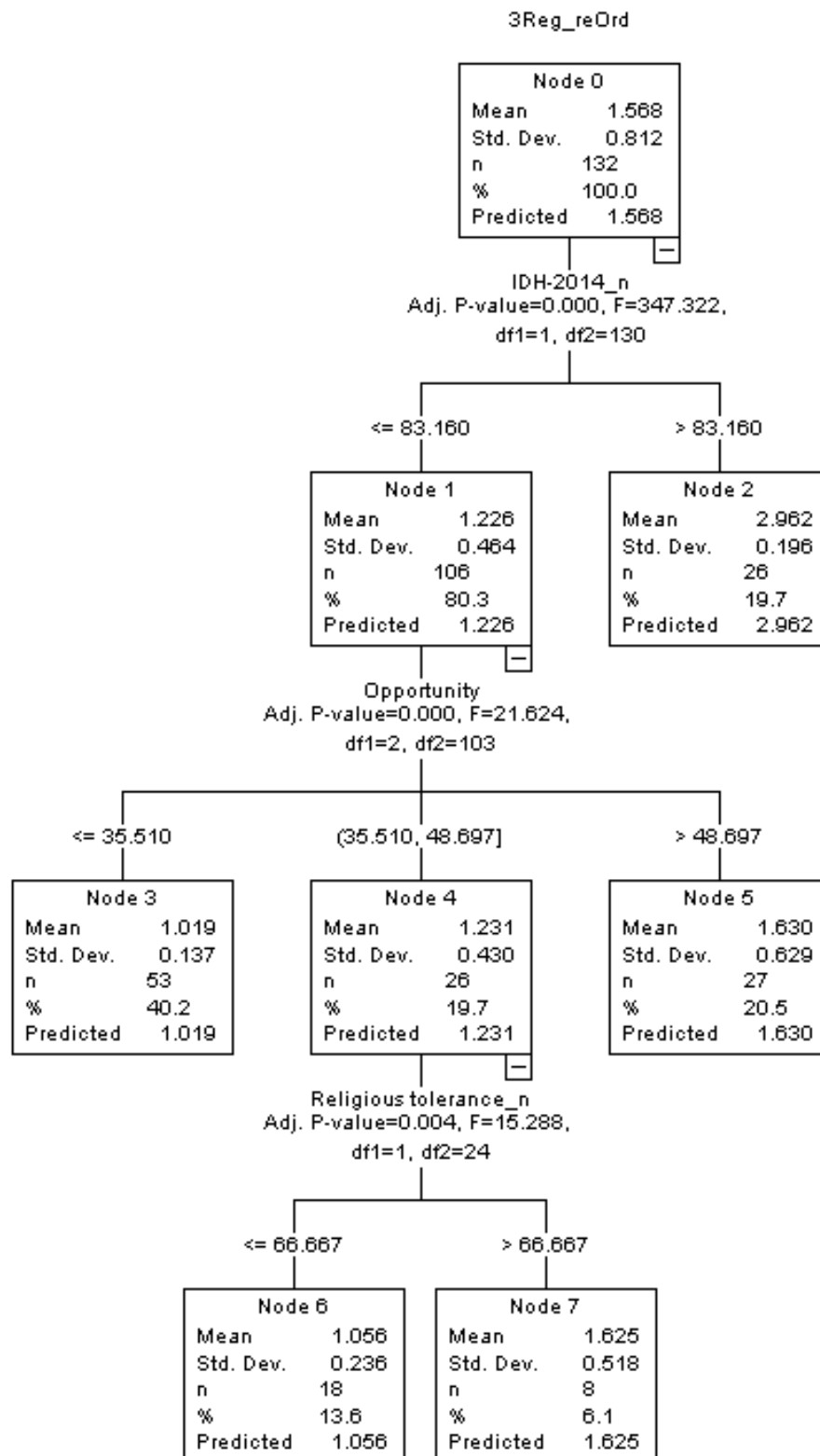
Classification Tree**Notes**

Output Created	22-May-2016 16:33:04
Comments	
Input	Data
	C:\Users\Arvores de Classificacao.sav

	Active Dataset	DataSet1	
	Filter	<none>	
	Weight	<none>	
	Split File	<none>	
	N of Rows in Working Data File		132
Missing Value Handling	Definition of Missing	Handling of user-defined missing values of nominal independent variables depends on the growing method.	
	Cases Used	Only cases with valid data for the dependent variable and some or all independent variables are used in computing any statistics.	
Syntax		<pre> TREE @3Reg_reOrd [s] BY Opportunity_n [s] Waterwithdrawals_np [s] Toleranceforimmigrants_n [s] Religioustolerance_n [s] IDH2014_n [s] Governança_n [s] EPI2014_n [s] HPI_n [s] DemocracyIndex2014_n [s] /TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES NODEDEFS=YES SCALE=AUTO /PRINT MODELSUMMARY RISK /GAIN SUMMARYTABLE=YES TYPE=[NODE] SORT=DESCENDING CUMULATIVE=NO /METHOD TYPE=EXHAUSTIVECHAID /GROWTHLIMIT MAXDEPTH=5 MINPARENTSIZE=10 MINCHILDSIZE=5 /VALIDATION TYPE=NONE OUTPUT=BOTHSAMPLES /CHAID ALPHASPLIT=0.05 SPLITMERGED=NO ADJUST=BONFERRONI INTERVALS=10. </pre>	
Resources	Processor Time		00:00:00.562
	Elapsed Time		00:00:01.556
Files Saved	Rules File		

Model Summary

Specifications	Growing Method	EXHAUSTIVE CHAID	
	Dependent Variable		3Reg_reOrd
	Independent Variables	Opportunity, Water withdrawals_np, Tolerance for immigrants_n, Religious tolerance_n, IDH-2014_n, Governance, EPI-2014_n, HPI, Democracy Index-2014_n	
	Validation	None	
	Maximum Tree Depth		5
	Minimum Cases in Parent Node		10
	Minimum Cases in Child Node		5
	Results	Independent Variables Included	IDH-2014_n, Opportunity, Religious tolerance_n
Number of Nodes			8
Number of Terminal Nodes			5
Depth			3



Gain Summary for Nodes

Node	N	Percent	Mean
2	26	19.7%	2.96
5	27	20.5%	1.63
7	8	6.1%	1.63
6	18	13.6%	1.06
3	53	40.2%	1.02

Growing Method: EXHAUSTIVE CHAID

Dependent Variable: 3Reg_reOrd

Risk

Estimate	Std. Error
.114	.025

Growing Method:

EXHAUSTIVE CHAID

Dependent Variable:

3Reg_reOrd

Método: CRT

* Decision Tree.

```

TREE @3Reg_reOrd [s] BY Opportunity_n [s] Waterwithdrawals_np [s] Toleranceforimmigrants_n [s]
Religioustolerance_n [s] IDH2014_n [s] Governança_n [s] EPI2014_n [s] HPI_n [s] DemocracyIndex2014_n [s]
/TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES NODEDEFS=YES
SCALE=AUTO
/PRINT MODELSUMMARY RISK
/GAIN SUMMARYTABLE=YES TYPE=[NODE] SORT=DESCENDING CUMULATIVE=NO
/METHOD TYPE=CRT MAXSURROGATES=AUTO PRUNE=NONE
/GROWTHLIMIT MAXDEPTH=5 MINPARENTSIZE=10 MINCHILDSIZE=5
/VALIDATION TYPE=NONE OUTPUT=BOTHSAMPLES
/CRT MINIMPROVEMENT=0.0001.

```

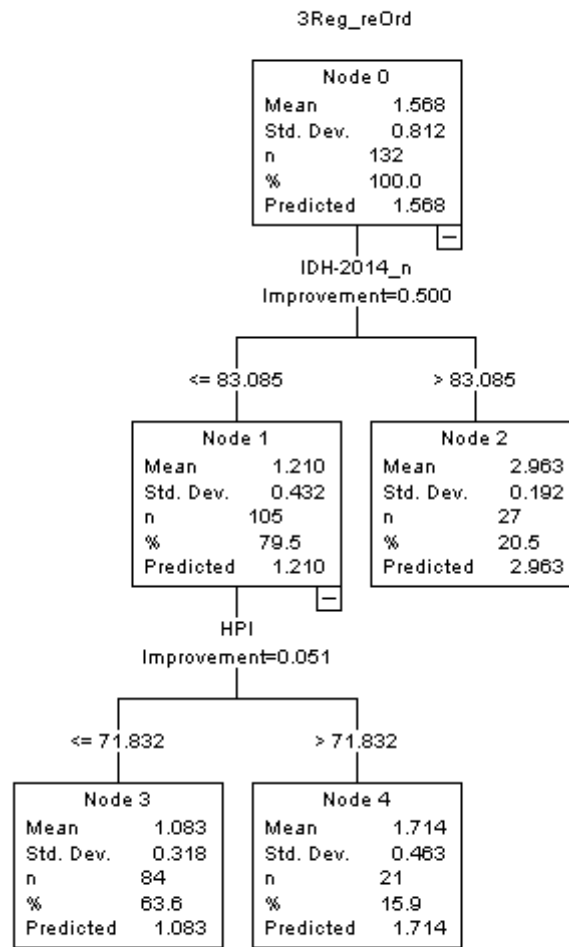
Classification Tree**Notes**

Output Created	22-May-2016 16:33:27
Comments	
Input	Data
	C:\Users\Arvores de Classificacao.sav

	Active Dataset	DataSet1	
	Filter	<none>	
	Weight	<none>	
	Split File	<none>	
	N of Rows in Working Data		132
	File		
Missing Value Handling	Definition of Missing	Handling of user-defined missing values of nominal independent variables depends on the growing method.	
	Cases Used	Only cases with valid data for the dependent variable and some or all independent variables are used in computing any statistics.	
Syntax		<pre> TREE @3Reg_reOrd [s] BY Opportunity_n [s] Waterwithdrawals_np [s] Toleranceforimmigrants_n [s] Religioustolerance_n [s] IDH2014_n [s] Governança_n [s] EPI2014_n [s] HPI_n [s] DemocracyIndex2014_n [s] /TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES NODEDEFS=YES SCALE=AUTO /PRINT MODELSUMMARY RISK /GAIN SUMMARYTABLE=YES TYPE=[NODE] SORT=DESCENDING CUMULATIVE=NO /METHOD TYPE=CRT MAXSURROGATES=AUTO PRUNE=NONE /GROWTHLIMIT MAXDEPTH=5 MINPARENTSIZE=10 MINCHILDSIZE=5 /VALIDATION TYPE=NONE OUTPUT=BOTHSAMPLES /CRT MINIMPROVEMENT=0.0001. </pre>	
Resources	Processor Time		00:00:00.624
	Elapsed Time		00:00:01.442
Files Saved	Rules File		

Model Summary

Specifications	Growing Method	CRT	
	Dependent Variable		3Reg_reOrd
	Independent Variables	Opportunity, Water withdrawals_np, Tolerance for immigrants_n, Religious tolerance_n, IDH-2014_n, Governance, EPI-2014_n, HPI, Democracy Index-2014_n	
	Validation	None	
	Maximum Tree Depth		5
	Minimum Cases in Parent Node		10
	Minimum Cases in Child Node		5
	Results	Independent Variables Included	IDH-2014_n, Opportunity, Governance, Democracy Index-2014_n, EPI-2014_n, Tolerance for immigrants_n, HPI
Number of Nodes			5
Number of Terminal Nodes			3
Depth			2



Gain Summary for Nodes

Node	N	Percent	Mean
2	27	20.5%	2.96
4	21	15.9%	1.71
3	84	63.6%	1.08

Growing Method: CRT

Dependent Variable: 3Reg_reOrd

Risk	
Estimate	Std. Error
.104	.032

Growing Method: CRT

Dependent Variable:

3Reg_reOrd

Analisando os resultados acima, vemos que o método CHAID apresenta o melhor resultado no crescimento das árvores (ramificação), resultando em 15 nós para a variável “ClusterM_reOrd”, e 10 nós para “3Reg_reOrd”.

11.5. CONSIDERAÇÕES

As análises discriminantes efetuadas no *software* Minitab, apresentaram maior grau de acerto que no SPSS, conforme mostra tabela comparativa na sessão anterior.

Com relação às árvores de classificação, o método CHAID apresentou o melhor resultado no crescimento das árvores (ramificação), resultando em 15 nós para a variável “ClusterM_reOrd”, e 10 nós para “3Reg_reOrd”, conforme resultados nas sessões 3.2.1 e 3.2.2.

12. ANÁLISE DE CORRESPONDÊNCIA

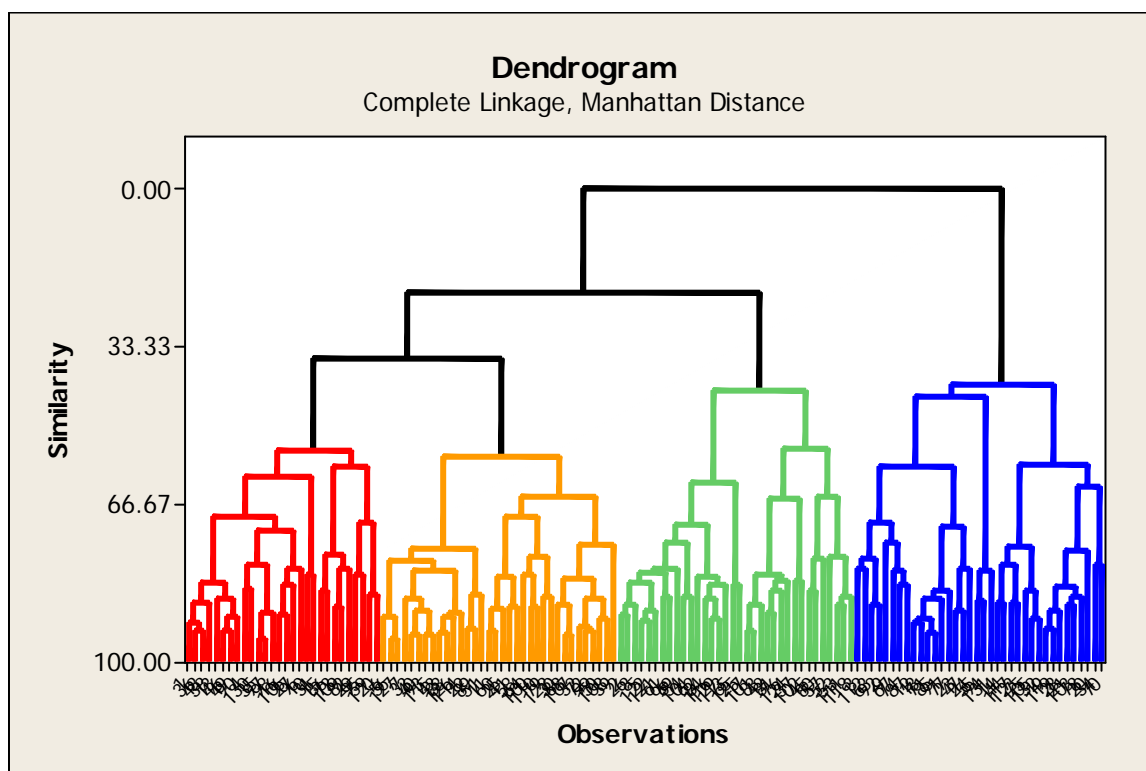
O objetivo de realizar análises de correspondência é obter graficamente a distribuição dos resultados das variáveis e analisar o grau de relacionamento entre as variáveis de interesse entre si, com as regiões assignadas, e com os agrupamentos de *clusters* selecionados. As análises serão efetuadas sobre as médias e os desvios-padrão de cada variável selecionada. Adicionalmente serão apresentados os coeficientes de determinação (r-quadrado) para auxiliar no entendimento da representatividade de cada variável e seu poder explicativo no modelo.

Obs: Conforme estabelecido, os países listados nesta tabela que não pertençam às regiões AIBER e AVECO, serão agrupados e classificados como OTHERS nas análises e resultados das sessões seguintes.

12.1. TRATAMENTO DOS DADOS, ANÁLISES DE CORRESPONDÊNCIA, AGRUPAMENTOS E RELACIONAMENTOS

Nesta sessão serão realizadas análises de correspondência dos dados dimensionadores do eixo temático Biodiversidade – Capital Natural e Social. Para apoiar estas análises, serão apresentados também os resultados em forma de gráfico radar relacionando os indicadores (variáveis) com as regiões e os agrupamentos de *clusters* pelas médias, dispersão (desvios-padrão) e grau de explicação dos modelos (r-quadrado).

As variáveis de trabalho para este estudo, conforme detalhadas na sessão anterior (quadro 1), são: EPI-2014, Governança, Democracy Index-2014, Opportunity, IDH-2014, Tolerance for immigrants, Water withdrawals, HPI (Happy Planet Index) e Religious tolerance. Além destas 9 (nove) variáveis, extraídas da base de dados deste estudo, 2 (duas) variáveis adicionais foram criadas para auxiliar nas análises. A variável “**3Reg_reOrd**”, que representa as 3 (três) regiões AIBER, AVECO e OTHERS ordenadas pelos resultados de uma ANOVA (análise de variância), classificando os resultados dos mais baixos (piores) para os mais altos (melhores). E a variável “**ClusterM_reOrd**”, que é resultado de 4 (quatro) agrupamentos por similaridade, através do método *Complete* e medida de distância *Manhattan*, classificados em ordem crescente de média.



12.2 Análises de correspondência

Nesta sessão será executada a função de Análise de correspondência (*simple correspondence analysis*) da ferramenta Minitab, relacionando primeiramente as regiões e os agrupamentos (*Row Plot*), para em seguida mostrar os relacionamentos das regiões, agrupamentos e as 9 (nove) variáveis deste estudo (*Symmetric Plot*).

A análise de correspondência é um método de análise fatorial para variáveis categóricas. Basicamente, converte uma tabela de dados não negativos de duas ou múltiplas entradas em um tipo de representação gráfica em que as linhas e as colunas são simultaneamente representadas em dimensão reduzida, isto é, por pontos no gráfico. Este método permite mostrar como as variáveis dispostas em linhas e colunas estão relacionadas e não somente se a relação existe.

As análises foram efetuadas para cruzamento e conhecer a correspondência, de média e desvio padrão, entre as 9 variáveis selecionadas, as 3 regiões e os 4 clusters.

12.2.1 Análise de Correspondência das Médias

Nas análises a seguir, algumas variáveis tiveram seus nomes reduzidos para melhor acomodação nos gráficos e tabelas. A variável *Opportunity*, tornou-se “opp”, *Water withdrawals* = “waterw”, *Tolerance for immigrants* = “immi”, *Religious tolerance* = “relig”, IDH = “idh”, *Governance* = “gov”, EPI = “epi”, HPI = “hpi” e *Democracy* = “democ”.

Simple Correspondence Analysis: opp, water, immi, relig, idh, gov, epi, hpi, democ

Contingency Table

	opp	waterw	immig	relig	idh	gov	epi	hpi
others	32.450	60.640	45.710	51.590	47.760	33.360	48.770	41.790
aiber	55.280	62.780	61.970	84.130	65.670	42.480	71.050	72.360
aveco	79.940	62.440	62.860	64.200	91.890	83.250	90.290	50.580
g1	21.720	81.500	53.920	53.700	24.390	23.690	23.900	34.500
g2	50.490	60.290	56.120	88.100	61.470	38.840	68.190	61.400
g3	34.930	36.760	34.370	31.370	66.130	36.730	65.620	52.780
g4	78.290	65.470	63.460	69.610	88.960	80.980	88.980	48.250
Total	353.100	429.880	378.410	442.700	446.270	339.330	456.800	361.660

	democ	Total
others	40.110	402.180
aiber	59.980	575.700
aveco	82.480	667.930
g1	33.030	350.350
g2	57.260	542.160
g3	38.900	397.590
g4	80.590	664.590
Total	392.350	3600.500

Chi-Square Distances

	opp	waterw	immig	relig	idh	gov	epi	hpi	democ
others	1.239	3.318	0.280	0.093	0.088	0.545	0.100	0.048	0.315
aiber	0.025	0.516	0.035	2.516	0.453	2.556	0.054	3.652	0.121
aveco	3.182	3.756	0.767	3.913	1.001	6.547	0.363	4.064	1.291
g1	4.649	37.622	7.940	2.619	8.344	2.636	9.500	0.014	0.694
g2	0.135	0.305	0.013	6.895	0.488	2.940	0.005	0.885	0.056
g3	0.423	2.416	1.316	6.276	5.761	0.015	4.567	4.130	0.452
g4	2.639	2.427	0.584	1.793	0.527	5.373	0.258	5.130	0.921
Total	12.291	50.360	10.936	24.104	16.662	20.611	14.847	17.923	3.851

	Total
others	6.025
aiber	9.929
aveco	24.884
g1	74.018
g2	11.722
g3	25.357
g4	19.653
Total	171.586

Relative Inertias

	opp	waterw	immig	relig	idh	gov	epi	hpi	democ	Total
others	0.007	0.019	0.002	0.001	0.001	0.003	0.001	0.000	0.002	0.035
aiber	0.000	0.003	0.000	0.015	0.003	0.015	0.000	0.021	0.001	0.058
aveco	0.019	0.022	0.004	0.023	0.006	0.038	0.002	0.024	0.008	0.145
g1	0.027	0.219	0.046	0.015	0.049	0.015	0.055	0.000	0.004	0.431
g2	0.001	0.002	0.000	0.040	0.003	0.017	0.000	0.005	0.000	0.068
g3	0.002	0.014	0.008	0.037	0.034	0.000	0.027	0.024	0.003	0.148
g4	0.015	0.014	0.003	0.010	0.003	0.031	0.002	0.030	0.005	0.115
Total	0.072	0.293	0.064	0.140	0.097	0.120	0.087	0.104	0.022	1.000

Analysis of Contingency Table

Axis	Inertia	Proportion	Cumulative	Histogram
1	0.0331	0.6951	0.6951	*****
2	0.0097	0.2037	0.8987	*****
3	0.0045	0.0949	0.9937	****
4	0.0003	0.0063	0.9999	
5	0.0000	0.0001	1.0000	
Total	0.0477			

Row Contributions

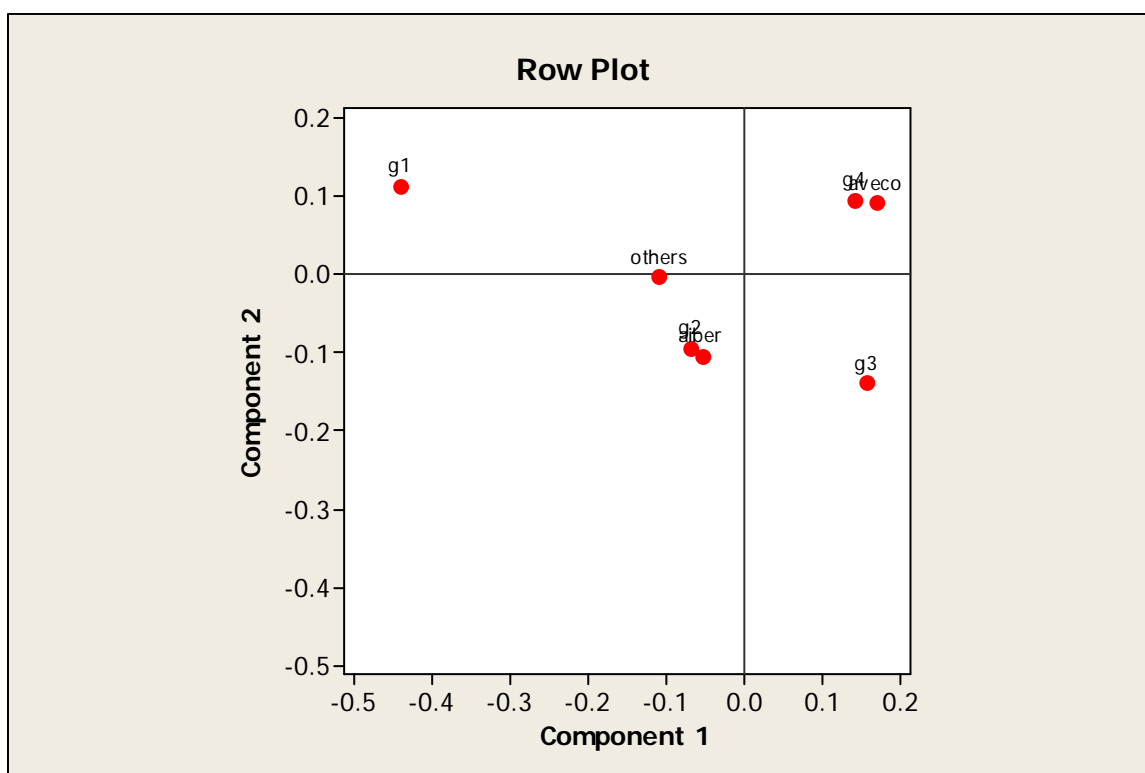
ID	Name	Qual	Mass	Inert	Coord	Corr	Contr	Coord	Corr	Contr
1	others	0.806	0.112	0.035	-0.110	0.806	0.041	-0.002	0.000	0.000
2	aiber	0.802	0.160	0.058	-0.055	0.176	0.015	-0.104	0.625	0.178
3	aveco	0.996	0.186	0.145	0.169	0.765	0.160	0.093	0.231	0.164
4	g1	0.984	0.097	0.431	-0.441	0.921	0.572	0.115	0.063	0.134
5	g2	0.625	0.151	0.068	-0.070	0.224	0.022	-0.093	0.401	0.134
6	g3	0.675	0.110	0.148	0.155	0.377	0.080	-0.138	0.298	0.216
7	g4	0.983	0.185	0.115	0.141	0.674	0.111	0.096	0.309	0.174

Column Contributions

ID	Name	Qual	Mass	Inert	Coord	Corr	Contr	Coord	Corr	Contr
1	opp	0.834	0.098	0.072	0.160	0.731	0.075	0.060	0.103	0.036
2	waterw	0.938	0.119	0.293	-0.317	0.857	0.362	0.097	0.081	0.117
3	immig	0.986	0.105	0.064	-0.163	0.920	0.084	0.044	0.066	0.021
4	relig	0.667	0.123	0.140	-0.181	0.603	0.122	-0.059	0.064	0.044
5	idh	0.895	0.124	0.097	0.178	0.844	0.118	-0.044	0.051	0.024
6	gov	0.997	0.094	0.120	0.195	0.624	0.108	0.151	0.374	0.220
7	epi	0.960	0.127	0.087	0.161	0.801	0.100	-0.072	0.159	0.068
8	hpi	0.945	0.100	0.104	-0.069	0.096	0.014	-0.205	0.849	0.435

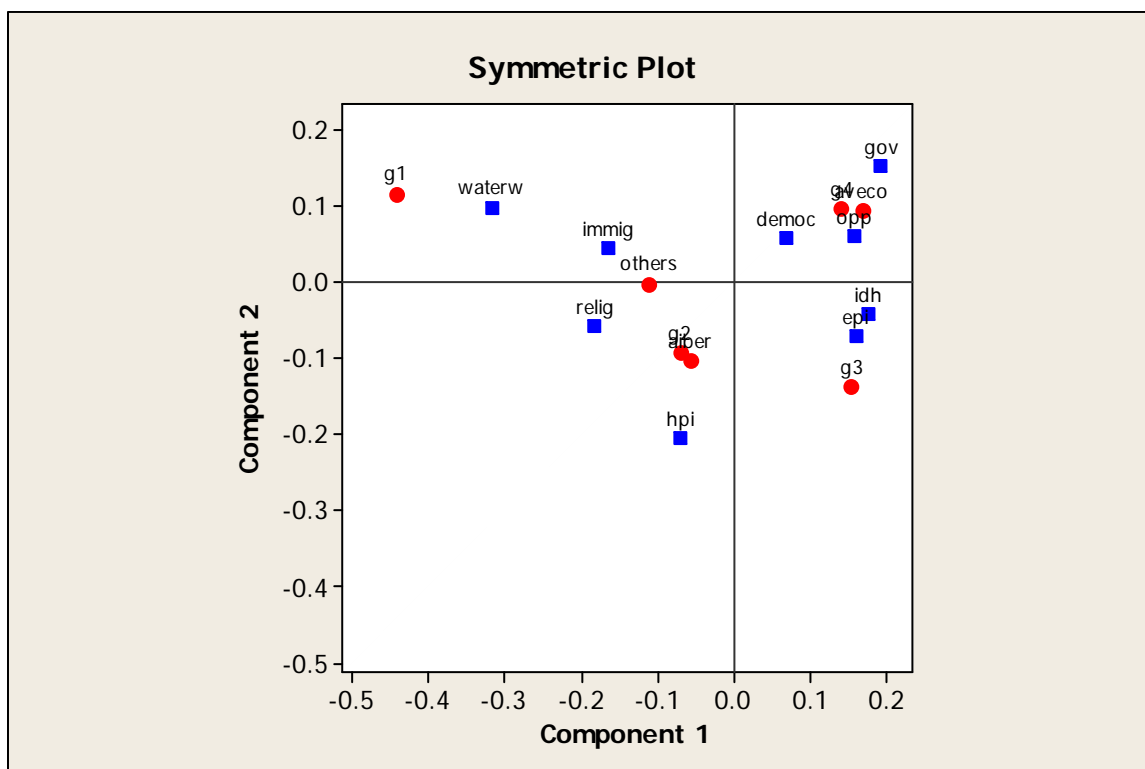
9 democ 0.830 0.109 0.022 0.071 0.512 0.017 0.056 0.318 0.035

Gráfico das médias, relacionando as 3 regiões e os 4 agrupamentos - Row Plot



O gráfico acima mostra a grande proximidade entre a região AVECO e o agrupamento G4, denotando forte relação entre estes dois grupos. Situação similar acontece entre AIBER e G2, apesar de não figurar positivamente no gráfico. E por fim, a região OTHERS que aparece num ponto médio entre G1 e G3. As proximidades e relações mostradas no gráfico acima, era de certa forma esperada já que quase a totalidade dos países de AVECO, foram categorizados no agrupamento G4, o mesmo com AIBER em relação ao G2 e os países da região OTHERS que foram distribuídos entre os G1 e G3, e figura num ponto médio entre estes dois, tanto no eixo horizontal, como no vertical.

Gráfico das médias, relacionando as 3 regiões, os 4 agrupamentos e as 9 variáveis - Symmetric Plot



Quando adicionamos as variáveis no gráfico, AIBER se destaca pela proximidade com os indicadores relig (tolerância às diferentes crenças religiosas) e hpi (índice de felicidade do planeta). Enquanto a região AVECO demonstra forte relação com os indicadores opp (oportunidades), democ (índice de democracia) e gov (governança), no quadrante positivo do gráfico. Pela grande dispersão de OTHERS no gráfico, os países desta região classificados como G1, se destacam negativamente pelo indicador waterw (retirada de água), enquanto G3, positivamente com idh (índice de desenvolvimento humano) e epi (índice de desempenho ambiental) em relação ao *component 1*.

12.2.2 Análise de Correspondência dos Desvios-Padrão

Nesta sessão, o processo será repetido utilizando as mesmas variáveis e metodologias da sessão anterior, porém utilizando os resultados de seus desvios-padrão para evidenciar distanciamentos em relação às médias. Quanto maior o desvio padrão, maior a dispersão e mais afastados da média estarão os eventos extremos.

Simple Correspondence Analysis: opp, water, immi, relig, idh, gov, epi, hpi, de

Contingency Table

	opp	waterw	immig	relig	idh	gov	epi	hpi
others	15.710	33.860	20.910	33.300	23.840	16.350	26.000	20.680
aiber	14.520	25.680	16.250	22.650	12.450	18.820	11.910	16.250
aveco	14.090	26.280	24.270	27.620	4.780	13.380	8.040	12.840
g1	11.650	27.200	20.380	35.880	13.480	12.090	15.370	19.040
g2	9.760	28.150	15.090	18.620	9.980	12.490	9.080	28.320
g3	10.380	27.050	18.040	21.620	12.710	14.030	14.190	16.560
g4	13.390	24.120	23.630	22.270	7.150	12.730	9.540	12.500
Total	89.500	192.340	138.570	181.960	84.390	99.890	94.130	126.190

democ Total

others	20.430	211.080
aiber	13.590	152.120
aveco	8.870	140.170
g1	16.310	171.400
g2	12.650	144.140
g3	21.270	155.850
g4	9.160	134.490
Total	102.280	1109.250

Chi-Square Distances

	opp	waterw	immig	relig	idh	gov	epi	hpi	democ	Total
others	0.102	0.205	1.130	0.051	3.771	0.372	3.652	0.463	0.048	9.793
aiber	0.411	0.018	0.399	0.213	0.066	1.915	0.077	0.064	0.014	3.177
aveco	0.684	0.160	2.609	0.931	3.247	0.045	1.249	0.605	1.272	10.803
g1	0.343	0.214	0.050	2.144	0.015	0.725	0.047	0.011	0.016	3.564
g2	0.301	0.399	0.472	1.068	0.089	0.019	0.812	8.669	0.031	11.858
g3	0.383	0.000	0.105	0.609	0.061	0.000	0.070	0.077	3.313	4.619
g4	0.594	0.027	2.776	0.002	0.928	0.032	0.307	0.512	0.847	6.026
Total	2.818	1.024	7.541	5.017	8.177	3.107	6.215	10.401	5.540	49.840

Relative Inertias

	opp	waterw	immig	relig	idh	gov	epi	hpi	democ	Total
others	0.002	0.004	0.023	0.001	0.076	0.007	0.073	0.009	0.001	0.196
aiber	0.008	0.000	0.008	0.004	0.001	0.038	0.002	0.001	0.000	0.064
aveco	0.014	0.003	0.052	0.019	0.065	0.001	0.025	0.012	0.026	0.217
g1	0.007	0.004	0.001	0.043	0.000	0.015	0.001	0.000	0.000	0.072
g2	0.006	0.008	0.009	0.021	0.002	0.000	0.016	0.174	0.001	0.238
g3	0.008	0.000	0.002	0.012	0.001	0.000	0.001	0.002	0.066	0.093
g4	0.012	0.001	0.056	0.000	0.019	0.001	0.006	0.010	0.017	0.121
Total	0.057	0.021	0.151	0.101	0.164	0.062	0.125	0.209	0.111	1.000

Analysis of Contingency Table

Axis	Inertia	Proportion	Cumulative	Histogram
1	0.0222	0.4932	0.4932	*****
2	0.0130	0.2903	0.7836	*****
3	0.0049	0.1098	0.8934	*****
4	0.0031	0.0696	0.9630	****
5	0.0016	0.0348	0.9978	**
6	0.0001	0.0022	1.0000	
Total	0.0449			

Row Contributions

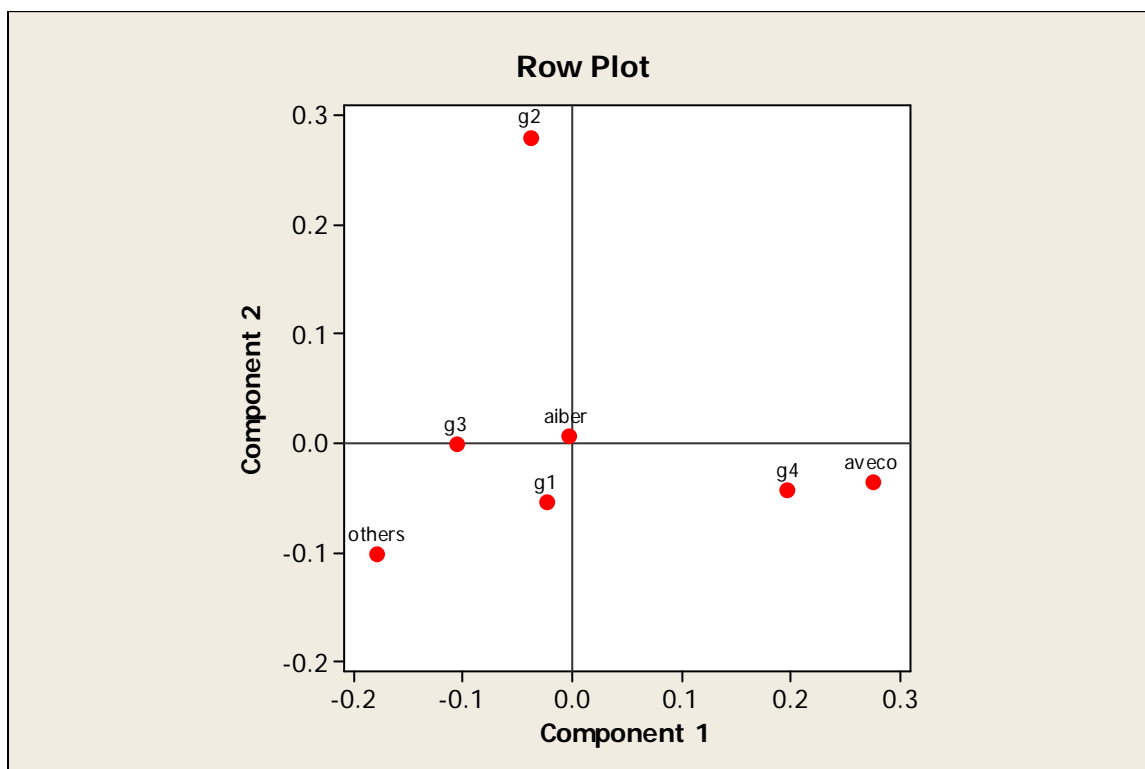
ID	Name	Qual	Mass	Inert	Coord	Corr	Component 1 Contr	Component 2 Coord	Corr	Contr
1	others	0.915	0.190	0.196	-0.179	0.693	0.276	-0.101	0.222	0.150
2	aiber	0.003	0.137	0.064	-0.004	0.001	0.000	0.007	0.002	0.001
3	aveco	0.990	0.126	0.217	0.274	0.973	0.428	-0.036	0.017	0.012
4	g1	0.168	0.155	0.072	-0.024	0.028	0.004	-0.054	0.140	0.035
5	g2	0.977	0.130	0.238	-0.039	0.018	0.009	0.281	0.958	0.785
6	g3	0.385	0.141	0.093	-0.107	0.385	0.072	-0.001	0.000	0.000
7	g4	0.901	0.121	0.121	0.196	0.861	0.211	-0.042	0.040	0.017

Column Contributions

ID	Name	Qual	Mass	Inert	Coord	Corr	Component 1 Contr	Component 2 Coord	Corr	Contr
1	opp	0.676	0.081	0.057	0.139	0.611	0.070	-0.045	0.065	0.013
2	waterw	0.843	0.173	0.021	0.039	0.291	0.012	0.054	0.552	0.039
3	immig	0.906	0.125	0.151	0.217	0.864	0.265	-0.048	0.042	0.022
4	relig	0.493	0.164	0.101	0.073	0.195	0.040	-0.091	0.298	0.103
5	idh	0.946	0.076	0.164	-0.293	0.885	0.294	-0.077	0.061	0.035

6	gov	0.135	0.090	0.062	0.060	0.114	0.014	0.026	0.022	0.005
7	epi	0.946	0.085	0.125	-0.205	0.635	0.161	-0.143	0.311	0.134
8	hpi	0.948	0.114	0.209	-0.060	0.044	0.018	0.273	0.904	0.650
9	democ	0.556	0.092	0.111	-0.174	0.556	0.125	-0.003	0.000	0.000

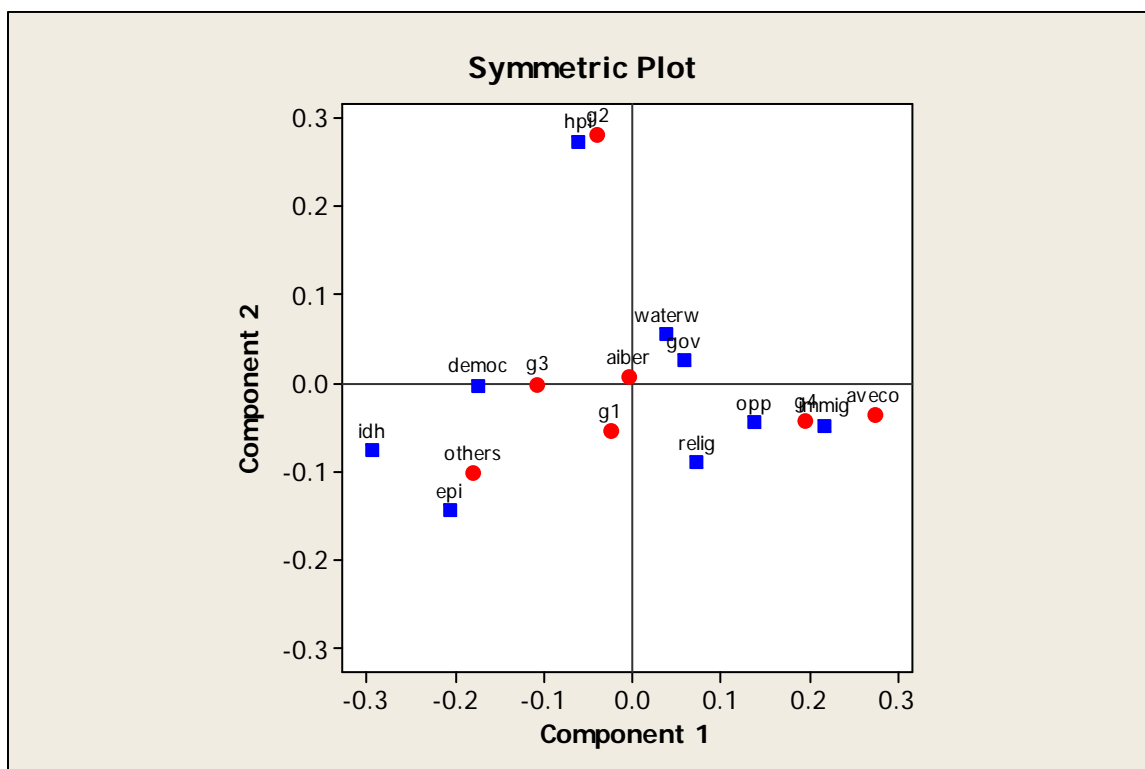
Gráfico dos desvios-padrão, relacionando as 3 regiões e os 4 agrupamentos - Row Plot



No gráfico acima destaca-se o agrupamento G2, com a maior dispersão (0.281) em relação ao eixo vertical (Component 2). O G2 é composto em sua quase totalidade por países do AIBER, que aparece no gráfico bem próximo de 0,0 para os dois eixos, provavelmente o modelo tenha capturado algum ruído que justifique este comportamento, e que aparentemente está relacionado ao fato de os países Espanha, Portugal, Chile e Uruguai de AIBER, estejam no G4. AVECO e OTHERS também apresentaram certa dispersão, mas não ficaram tão distantes de seus agrupamentos.

AIBER destacou-se positivamente por situar-se na parte mais central do gráfico, mostrando a menor dispersão em relação aos dois eixos (*component 1 e 2*)

Gráfico dos desvios-padrão, relacionando as 3 regiões, os 4 agrupamentos e as 9 variáveis - Symmetric Plot



Como vimos na análise do gráfico anterior, o agrupamento G2 é composto em sua maioria por países de AIBER, com exceção de Espanha, Portugal, Chile e Uruguai, e como vínhamos notando nos trabalhos anteriores, este grupo apresentava uma relação bem forte com o hpi (índice de felicidade do planeta), o que percebemos também neste gráfico. As outras variáveis também apresentaram certo nível de dispersão, em sua maioria no eixo horizontal (*component 1*), mas mantiveram certa proximidade com suas regiões e agrupamentos de *clusters*.

12.3 Agrupamentos e Relacionamentos

Nesta sessão serão apresentados gráficos do tipo radar e tabelas de resumo para mostrar como as variáveis se agrupam, se relacionam e seus graus de explicação.

Conforme mencionado na sessão anterior, algumas variáveis tiveram seus nomes reduzidos para melhor acomodação nos gráficos e tabelas. A variável *Opportunity*, tornou-se “opp”, *Water withdrawals* = “waterw”, *Tolerance for immigrants* = “immi”, *Religious tolerance* = “relig”, IDH = “idh”, *Governance* = “gov”, EPI = “epi”, HPI = “hpi” e *Democracy* = “democ”.

As análises foram efetuadas para conhecer a correspondência, de média e desvio padrão, entre as 9 variáveis selecionadas, as 3 regiões e os 4 clusters.

12.3.1 Tabela com as médias e gráficos radar

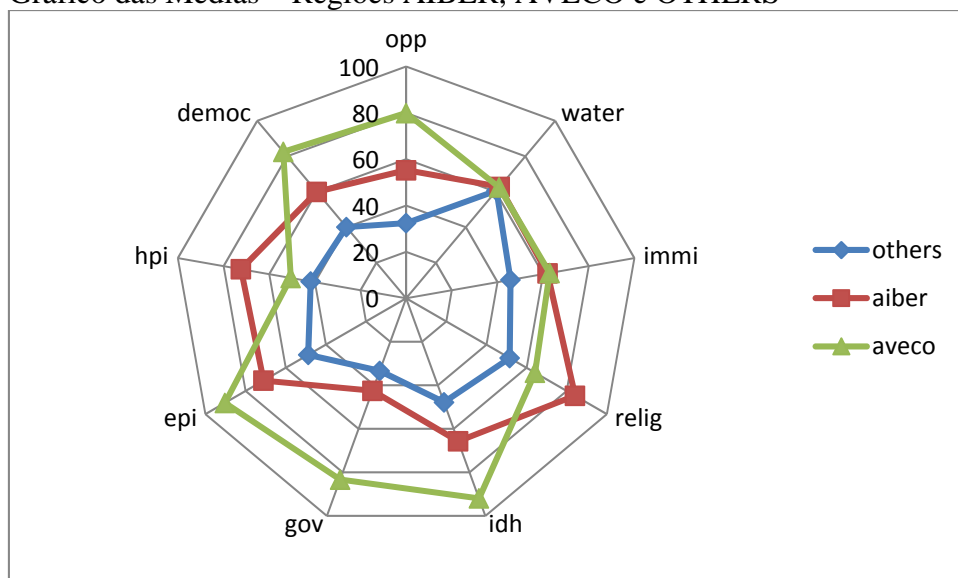
A tabela abaixo mostra as médias das variáveis contemplando as 3 regiões (others, aiber e aveco) e os 4 agrupamentos (clusters). E o coeficiente de determinação (r^2).

MÉDIA									
Mean	opp	waterw	immi	relig	idh	gov	epi	hpi	democ
others	32.45	60.64	45.71	51.59	47.76	33.36	48.77	41.79	40.11
aiber	55.28	62.78	61.97	84.13	65.67	42.48	71.05	72.36	59.98
aveco	79.94	62.44	62.86	64.20	91.89	83.25	90.29	50.58	82.48
R-Sq	61.81%	0.09%	13.12%	13.21%	44.45%	60.07%	38.44%	26.02%	48.68%
g1	21.72	81.50	53.92	53.70	24.39	23.69	23.90	34.50	33.03
g2	50.49	60.29	56.12	88.10	61.47	38.84	68.19	61.40	57.26
g3	34.93	36.76	34.37	31.37	66.13	36.73	65.62	52.78	38.90
g4	78.29	65.47	63.46	69.61	88.96	80.98	88.98	48.25	80.59
R-Sq	78.52%	28.32%	23.88%	38.72%	82.45%	75.06%	79.54%	20.27%	60.18%

Nesta tabela figuram em cores de destaque as variáveis que apresentaram as 3 (três) maiores médias (cor verde) e as 3 menores (cor laranja), e sua respectiva região e agrupamento. Mostra também em destaque as variáveis que apresentaram o maior r-quadrado por região e por agrupamento (cor azul) e o menor (cor vermelho).

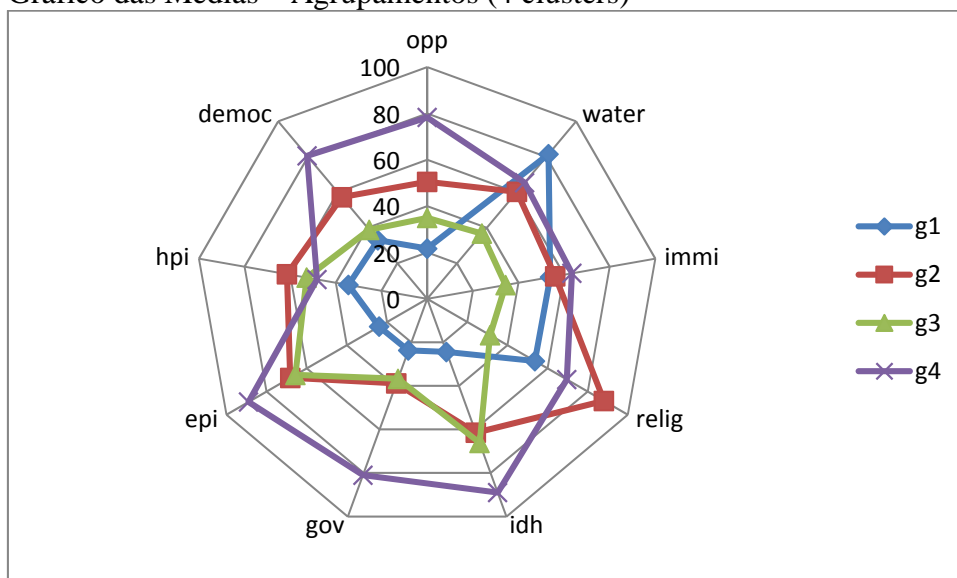
Pelos resultados acima, vemos que a região OTHERS destacou-se negativamente com as menores médias para as variáveis opp, gov e democ. Comportamento similar nos agrupamentos, onde G1 resultou nas menores médias para opp, gov e epi. No lado positivo, as maiores médias foram para as variáveis idh e epi para AVECO e relig para AIBER. Desempenho análogo nos agrupamentos, com destaque para idh e epi no G4, e relig para G2 (ver r-quadrado).

Gráfico das Médias – Regiões AIBER, AVECO e OTHERS



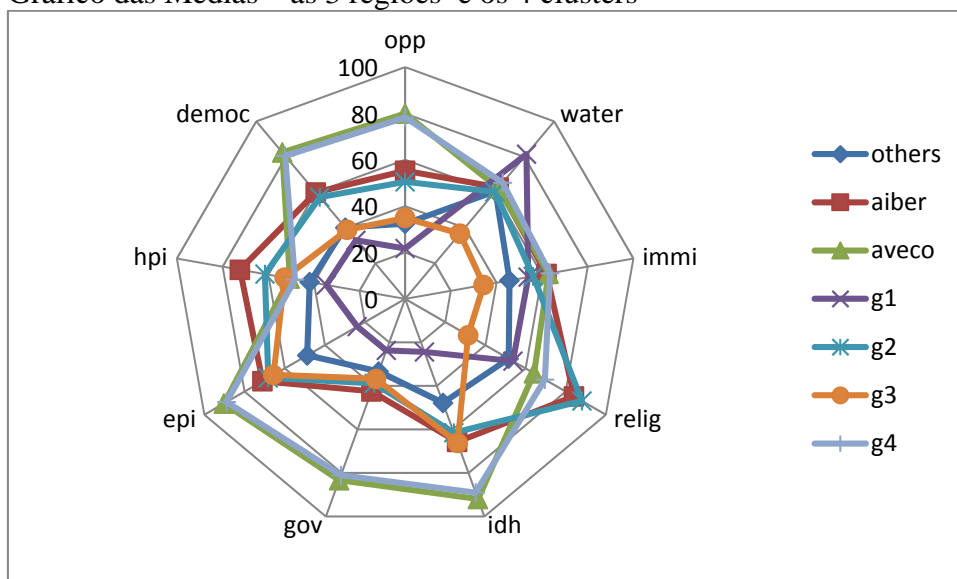
Os maiores destaques para AIBER neste gráfico estão relacionados às variáveis hpi e relig.

Gráfico das Médias – Agrupamentos (4 clusters)



O agrupamento G2, formado predominantemente por países AIBER se destaca positivamente pelo índice relig (*religious tolerance*). E o G1, positivamente pelo water (retirada de água/bombeamento).

Gráfico das Médias – as 3 regiões e os 4 clusters



Quando sobrepostos os gráficos, o indicador relig continua destacando AIBER e G2.

12.3.2 Tabela com os desvios-padrão e gráfico radar

A tabela e gráficos abaixo mostram os desvios-padrão das variáveis contemplando as 3 regiões (others, aiber e aveco) e os 4 agrupamentos (clusters). E o coeficiente de determinação (r-quadrado).

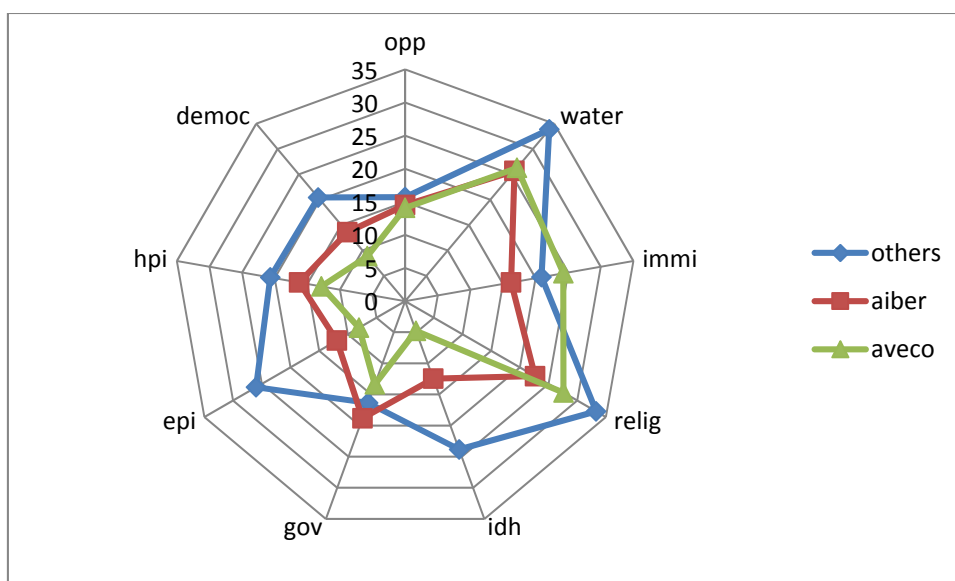
Desvio-Padrão									
StDev	opp	waterw	immi	relig	idh	gov	epi	hpi	democ

others	15.71	33.86	20.91	33.30	23.84	16.35	26.00	20.68	20.43
aiber	14.52	25.68	16.25	22.65	12.45	18.82	11.91	16.25	13.59
aveco	14.09	26.28	24.27	27.62	4.78	13.38	8.04	12.84	8.87
R-Sq	61.81%	0.09%	13.12%	13.21%	44.45%	60.07%	38.44%	26.02%	48.68%
g1	11.65	27.20	20.38	35.88	13.48	12.09	15.37	19.04	16.31
g2	9.76	28.15	15.09	18.62	9.98	12.49	9.08	28.32	12.65
g3	10.38	27.05	18.04	21.62	12.71	14.03	14.19	16.56	21.27
g4	13.39	24.12	23.63	22.27	7.15	12.73	9.54	12.50	9.16
R-Sq	78.52%	28.32%	23.88%	38.72%	82.45%	75.06%	79.54%	20.27%	60.18%

Nesta tabela figuram em cores de destaque as variáveis que apresentaram as 3 (três) maiores médias (cor verde) e as 3 menores (cor laranja), e sua respectiva região e agrupamento. Mostra também em destaque as variáveis que apresentaram o maior r-quadrado por região e por agrupamento (cor azul) e o menor (cor vermelho).

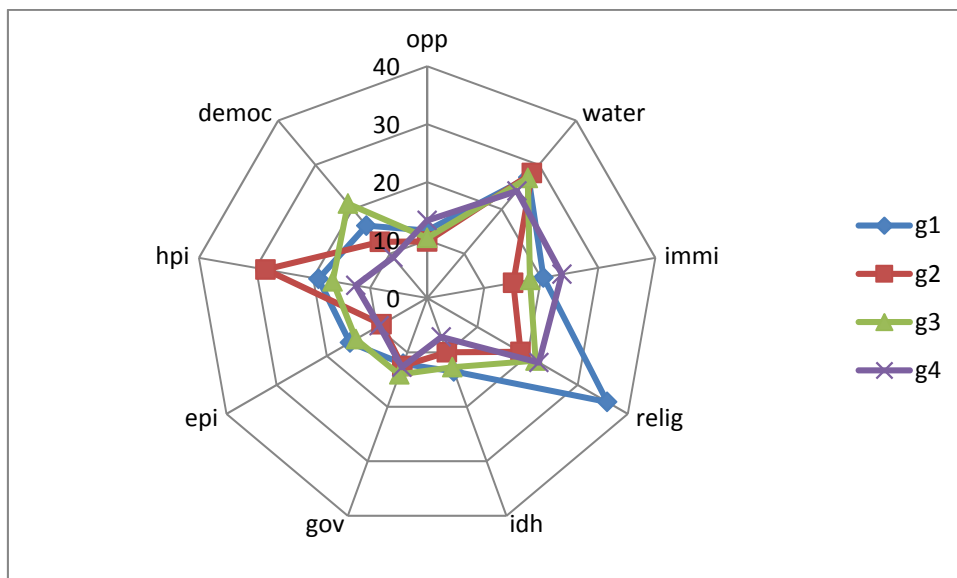
Os resultados acima mostram os maiores desvios-padrão para waterw e relig para a região OTHERS e relig para AVECO (ver r² muito baixo para as três). E waterw e hpi para G2, e relig para G1. Nos destaques positivos, ou seja, com os menores valores temos idh, epi e democ para AVECO. E nos agrupamentos, idh e democ para G4 e epi para G2.

Desvio-padrão – 3 regiões



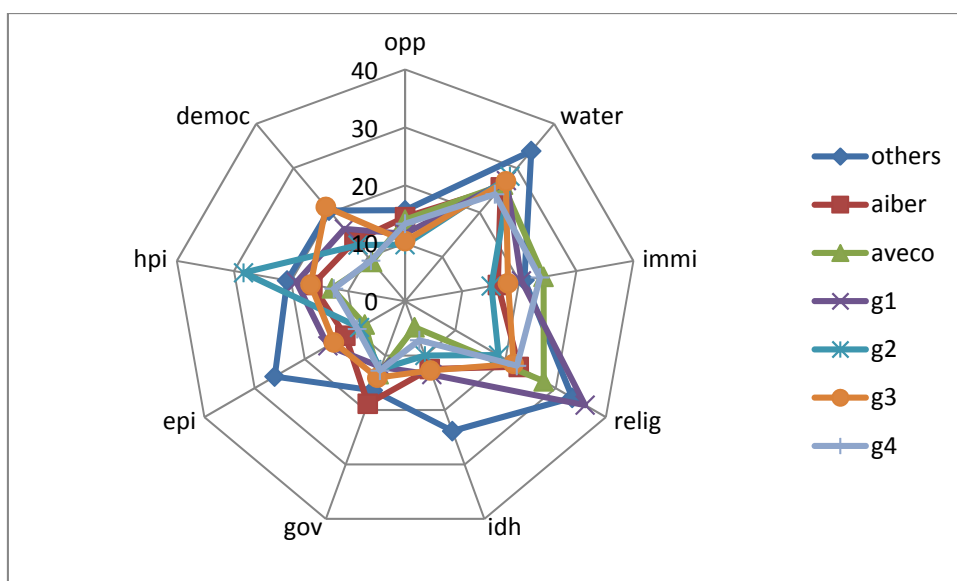
Em AIBER, waterw e relig apresentaram os maiores afastamentos de suas médias.

Desvio-padrão – 4 clusters



Similar ao gráfico da sessão anterior, hpi e waterw apresentaram as maiores dispersões no G2.

Desvio-padrão – as 3 regiões e os 4 clusters



Nesta sobreposição, AIBER destaca-se positivamente em epi. E G2 negativamente em hpi.

12.4. CONSIDERAÇÕES

Em média os resultados da região AVECO estiveram muito próximos aos do agrupamento G4, que se destacaram positivamente em quase todos os indicadores. Similarmente AIBER se relacionando com G2, figurando numa faixa intermediária nos resultados. E OTHERS com países nos agrupamentos G1 e G3, apresentando resultados médios e baixos.

Houve destaque positivo para AIBER em relação aos indicadores relig (tolerância às crenças religiosas) e hpi (índice de felicidade do planeta), mesmo quando a região foi

analisada como agrupamento G2, ou seja, sem os países Espanha, Portugal, Chile e Uruguai. Apesar destes resultados, AIBER (e G2) ainda tem um difícil caminho a trilhar, haja vista os baixos resultados em índices importantes ao desenvolvimento como gov, idh, democ, opp e epi.

13. Análises dos países da região Ibero-Americana (AIBER)

Este sessão se dedicará unicamente a analisar os países da região **Ibero-Americana (AIBER)**. Para tanto serão efetuadas análises de correspondência, para obter graficamente a distribuição dos resultados das variáveis do eixo temático **Biodiversidade - Capital Natural e Social** e analisar o grau de relacionamento entre as variáveis de interesse. Análises de variância (ANOVA), regressões, de agrupamentos (clusters) e de componentes principais também serão efetuadas. E por fim, ao se identificar as variáveis de maior relevância na composição do modelo, será criado um indicador de performance, através do qual os países serão agrupados e ordenados. E um mapa indicando suas posições no ranking através das cores para destaque.

13.1. ENTENDENDO OS DADOS

Os dados desta análise compreende os 21 países pertencentes à região Ibero-Americana (AIBER). Estes dados são representados pelos indicadores sintéticos e analíticos de maior relevância para permitir análises e conclusões sobre o tema **Biodiversidade - Capital Natural e Social** e auxiliar os envolvidos no Projeto ORIBER – Observatório RIBER.

13.1.2 As variáveis do tema Biodiversidade - Capital Natural e Social

Este tema de pesquisa apresenta um total de 33 (trinta e três) variáveis, sendo três categóricas – País, Country code e REGIÃO e trinta variáveis quantitativas, porém neste trabalho utilizaremos somente as variáveis identificadas como de alto poder explicativo, conforme mostra o Quadro 1.

Quadro 1. As variáveis de interesse para este trabalho

Variável	Significado	Tipo	Unidade de Medida
opp (Opportunity)	Indicador composto por: Direitos individuais, Liberdades individuais, Tolerância e inclusão e	Variável Quantitativa	Indicador nominal

	Acesso à educação superior		
waterw (Water withdrawals as a percent of resources)	Uso de água retirada como porcentagem dos recursos. Obedece uma escala de 0 a 5, sendo: 4–5: Extremely high stress (>80%), 3–4: High stress (40–80%), 2–3: Medium-high stress (20–40%), 1–2: Low-medium stress (10–20%) e 0–1: Low stress (<10%) (variável positivada)	Variável Quantitativa	Indicador nominal que varia de 0 a 5
immi (Tolerance for immigrants)	Variável que contabiliza respostas SIM (positivas) da população sobre a percepção do local onde vivem ser um bom lugar para imigrantes de outros países.	Variável Quantitativa	Percentual
relig (Religious tolerance)	Esta variável é a medida de 13 tipos de hostilidades causadas por indivíduos, organizações ou sociedade (assédio, abuso, intimidação,...). A escala varia de 1 (baixa tolerância) a 4 (alta tolerância).	Variável Quantitativa	Indicador nominal que varia de 1 a 4
idh (IDH-2014)	O Índice de Desenvolvimento Humano (IDH) mede o progresso de uma nação a partir de três dimensões: Saúde (vida longa e saudável; expectativa de vida), Educação (anos de escolaridade para adultos e crianças) e Renda (padrão de vida medido pela Renda Nacional Bruta – RNB).	Variável Quantitativa	Indicador nominal
gov (Governança)	Indicador que mede a capacidade de um país de eleger, monitorar e substituir seus governantes eleitos. Mede também nível de estabilidade política, medidas contra corrupção, entre outros.	Variável Quantitativa	Indicador nominal
epi (EPI-2014)	Índice de Performance Ambiental (<i>Environmental Performance Index</i>), composto por: Saúde Ambiental (saúde humana e dos ecossistemas), Impactos na Saúde (medido pela mortalidade infantil – idade entre 1 e 5 anos), Qualidade do Ar (poluição do ar, partículas sólidas e seu impacto na saúde) e Nível de Água tratada e Sanitização adequada.	Variável Quantitativa	Indicador nominal
hpi (HPI-Happy Planet Index)	HPI, Índice de Felicidade do Planeta, medida do Bem Estar Sustentável (Longevidade, Felicidade e Vida sustentável percebida pela população do país).	Variável Quantitativa	Indicador nominal
democ (Democracy Index – 2014)	A variável Índice de Democracia é formada por 5 categorias: Processo eleitoral e pluralismo, Liberdade civil, Funcionamento do governo, Participação política e Cultura política.	Variável Quantitativa	Indicador nominal

Fonte: O autor a partir dos dados da planilha estatística, do GPS p.16 e do site do Banco Mundial. A pontuação foi padronizada para uma escala de 0 a 100 e as variáveis foram positivadas.

13.1.3 Países analisados

Este trabalho consiste em fazer análises com os indicadores mencionados na sessão anterior abrangendo os 21 países da região Ibero-Americana, conforme o Quadro 2.

Quadro 2. Países contemplados neste estudo

Ord.	PAÍS	COD. PAÍS	REGIÃO
1	Argentina	ARG	AIBER
2	Bolivia	BOL	AIBER
3	Brazil	BRA	AIBER
4	Chile	CHL	AIBER
5	Colombia	COL	AIBER
6	Costa Rica	CRI	AIBER
7	Cuba	CUB	AIBER
8	Dominican Republic	DOM	AIBER
9	Ecuador	ECU	AIBER
10	El Salvador	SLV	AIBER
11	Guatemala	GTM	AIBER
12	Honduras	HND	AIBER
13	Mexico	MEX	AIBER
14	Nicaragua	NIC	AIBER
15	Panama	PAN	AIBER
16	Paraguay	PRY	AIBER
17	Peru	PER	AIBER
18	Portugal	PRT	AIBER
19	Spain	ESP	AIBER
20	Uruguay	URY	AIBER
21	Venezuela	VEN	AIBER

Obs: Conforme estabelecido, Cuba que estava originalmente classificado como um país da região do Caribe, passou a figurar como AIBER para as análises e resultados deste trabalho.

13.2. TRATAMENTO DOS DADOS, AGRUPAMENTOS E RELACIONAMENTOS

Nesta sessão serão realizadas análises de agrupamentos (*clusters*), regressões, análise da variância (ANOVA) e de componentes principais.

As variáveis de trabalho para este estudo, conforme detalhadas na sessão anterior (quadro 1), são: EPI-2014, Governança, Democracy Index-2014, Opportunity, IDH-2014, Tolerance for immigrants, Water withdrawals, HPI (Happy Planet Index) e Religious tolerance.

Tomando como base os resultados das componentes principais, será criado um indicador com as variáveis de maior poder de explicação do modelo e uma tabela de ranking,

ordenando os países pela sua performance e um mapa para mostrar graficamente estes resultados.

13.2.1 Análises de agrupamentos (*clusters*)

Nesta sessão será efetuada análises de agrupamentos a fim de evidenciar o grau de relacionamento das variáveis estudadas. Algumas variáveis tiveram seus nomes reduzidos para melhor acomodação nos gráficos e tabelas. A saber, a variável *Opportunity*, tornou-se “opp”, *Water withdrawals* = “waterw”, *Tolerance for immigrants* = “immi”, *Religious tolerance* = “relig”, IDH = “idh”, *Governance* = “gov”, EPI = “epi”, HPI = “hpi” e *Democracy* = “democ”.

Quanto a quantidade de *clusters*, foram testadas de 3 a 8 *clusters*, porém a que melhor agrupamento formava foi a de 4 *clusters*, portanto, a que será utilizada no transcórre das análises que se seguem.

Abaixo a análise de *clusters*, utilizando o método completo (*complete*), medida de distância pelo sistema *Manhattan*, e quantidade de *clusters* igual a 4 (quatro).

Cluster Analysis of Observations: opp, waterw, immi, relig, idh, gov, epi, ...

Manhattan Distance, Complete Linkage
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	20	91.1800	31.821	8 17	8	2
2	19	88.2816	42.277	11 12	11	2
3	18	82.8518	61.867	9 10	9	2
4	17	82.8070	62.029	11 14	11	3
5	16	80.2934	71.097	1 3	1	2
6	15	78.1834	78.710	18 19	18	2
7	14	75.3982	88.758	2 11	2	4
8	13	73.0828	97.112	2 9	2	6
9	12	68.6452	113.122	7 8	7	3
10	11	68.5229	113.563	4 18	4	3
11	10	67.7118	116.489	6 20	6	2
12	9	63.9665	130.001	1 21	1	3
13	8	63.0945	133.147	5 16	5	2
14	7	60.9999	140.704	2 15	2	7
15	6	52.9883	169.608	1 5	1	5
16	5	52.4140	171.680	4 6	4	5
17	4	50.6913	177.895	7 13	7	4
18	3	42.1679	208.646	1 2	1	12
19	2	27.4809	261.634	1 7	1	16
20	1	0.0000	360.779	1 4	1	21

Final Partition
Number of clusters: 4

Cluster	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	5	6175.88	34.5654	40.8627

Cluster2	7	4606.30	24.9371	36.1484
Cluster3	5	7585.58	38.4766	46.4003
Cluster4	4	5594.84	35.1330	54.6214

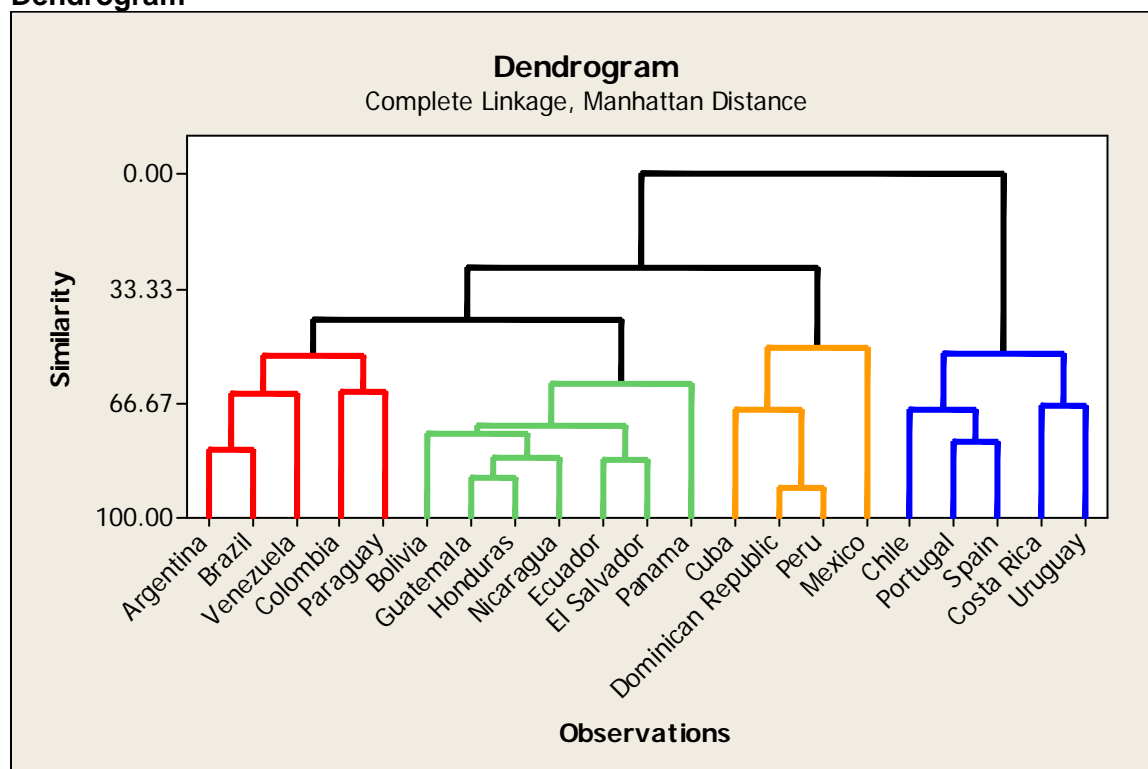
Cluster Centroids

Variable	Cluster1	Cluster2	Cluster3	Cluster4	Grand centroid
opp	53.4788	47.515	76.8383	44.1758	55.2807
waterw	75.7362	81.306	47.7046	33.0142	62.7811
immi	72.0548	52.446	77.5342	46.5753	61.9700
relig	60.0000	100.000	86.6667	83.3333	84.1270
idh	67.5320	54.346	79.0982	66.3773	65.6706
gov	30.7694	33.326	71.5665	36.8007	42.4840
epi	73.0720	60.660	83.5088	71.1553	71.0544
hpi	75.5122	77.386	61.2983	73.4319	72.3563
democ	58.4123	53.842	76.7536	51.6884	59.9752

Distances Between Cluster Centroids

	Cluster1	Cluster2	Cluster3	Cluster4
Cluster1	0.0000	49.0913	67.2170	56.5356
Cluster2	49.0913	0.0000	78.5934	54.2470
Cluster3	67.2170	78.5934	0.0000	67.4596
Cluster4	56.5356	54.2470	67.4596	0.0000

Dendrogram



Este dendrograma mostra os agrupamentos feitos pelos países por similaridade. Os países que figuram na cor azul, são os de melhores resultados na análise. Os de cor vermelha e laranja, apresentaram resultados intermediários. E os de cor verde apresentaram os valores mais baixos nesta análise.

Pelas conexões das linhas pretas, percebe-se de certa forma relação entre os grupos de cores vermelha e verde, e estes com o de cor laranja.

Ao compararmos a distância entre os coeficientes de correlação (cluster de variáveis), temos os seguintes resultados:

Cluster Analysis of Variables: opp, waterw, immi, relig, idh, gov, epi, democ,...

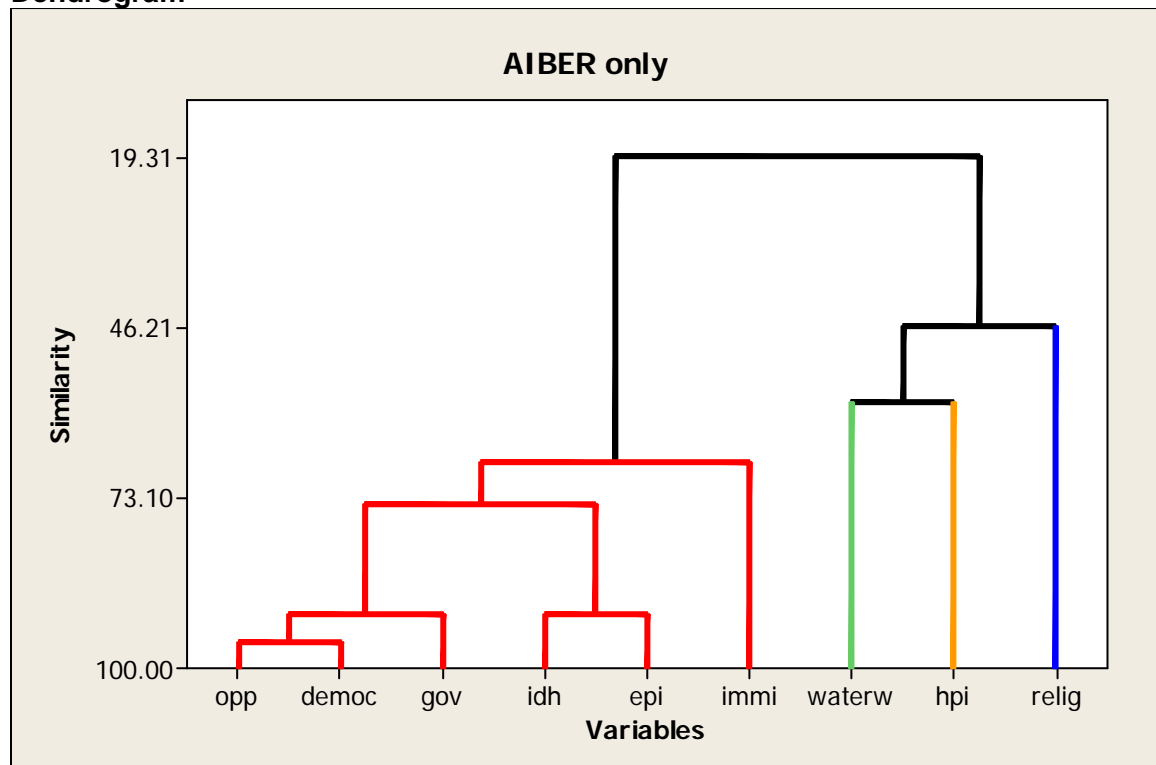
Correlation Coefficient Distance, Complete Linkage
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	8	95.9319	0.08136	9	1	2
2	7	91.5211	0.16958	5	7	2
3	6	91.4732	0.17054	1	6	3
4	5	73.9669	0.52066	1	5	5
5	4	67.3468	0.65306	1	3	6
6	3	57.8964	0.84207	2	8	2
7	2	46.1059	1.07788	2	4	3
8	1	19.3136	1.61373	1	2	9

Final Partition

```
Cluster 1
  opp immi idh gov epi democ
Cluster 2
  waterw
Cluster 3
  relig
Cluster 4
  hpi
```

Dendrogram



Pela similaridade encontrada na análise da ferramenta estatística, as 6 (seis) primeiras variáveis (cor vermelha), formaram um agrupamento, enquanto as demais não apresentaram similaridade suficiente para formar outros agrupamentos.

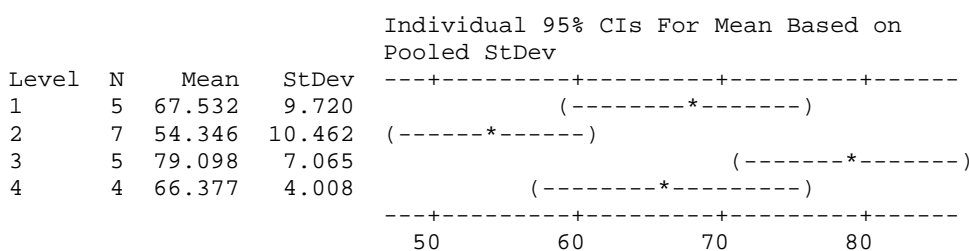
13.2.2 Análise de variância - ANOVA

Com os agrupamentos definidos e estabelecido cada país como membro de seu respectivo *cluster*, como próximo passo será executado a ANOVA para entender a variabilidade entre os grupos e auxiliar na classificação para as análises nas sessões posteriores.

One-way ANOVA: idh versus ClusterM

Source	DF	SS	MS	F	P
ClusterM	3	1818.5	606.2	8.04	0.001
Error	17	1282.5	75.4		
Total	20	3101.0			

S = 8.686 R-Sq = 58.64% R-Sq(adj) = 51.35%



Pooled StDev = 8.686

Na análise acima, vemos o que havia sido descrito na sessão anterior, análise de dendrograma, que o grupo 3, no gráfico representado pela cor azul, apresenta a maior média entre todos. Os grupos 1 e 4, com médias entre 66 e 67.5, bem próximas e quase se sobrepondo completamente, representados no gráfico da sessão anterior pelas cores vermelha e laranja, numa posição intermediária. E finalmente o grupo 2, com média de 54.346, os resultados mais baixos da amostra e representados no gráfico pela cor verde.

A coluna de ordenamento ficou com a seguinte característica de grupos:

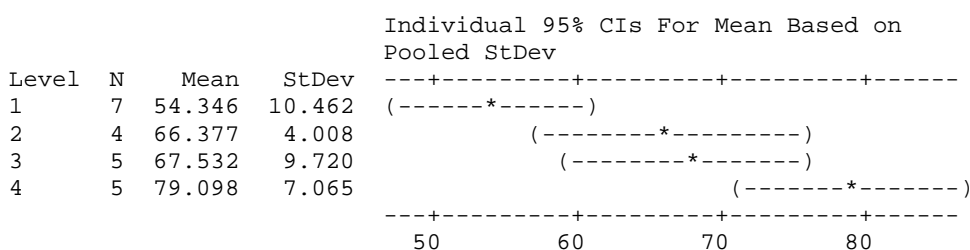
- De 2 para 1
- De 4 para 2
- De 1 para 3
- De 3 para 4

Feito o ordenamento (*cluster members*), é executado novamente a ANOVA com os seguintes resultados:

One-way ANOVA: idh versus ClusterM_reOrd

Source	DF	SS	MS	F	P
ClusterM_reOrd	3	1818.5	606.2	8.04	0.001
Error	17	1282.5	75.4		
Total	20	3101.0			

S = 8.686 R-Sq = 58.64% R-Sq(adj) = 51.35%



Pooled StDev = 8.686

Esta disposição reflete no grupo 1 com os resultados mais baixos, os grupos 2 e 3 os intermediários e o grupo 4, com os melhores.

Apesar de seguir com os 4 (quarto) clusters para as análises das próximas sessões, como indica os resultados da ferramenta Minitab, estamos alterando as classificações dos grupos 2 e 3, para que representem a categoria dos países com indicadores de nível “MÉDIO” devido sua proximidade de valores e sua quase sobreposição demonstrado no histograma da ANOVA (Análise de Variância).

Portanto, a nova classificação será:

De 1 para 1 - BAIXO

De 2 para 2 - MÉDIO

De 3 para 2 - MÉDIO

De 4 para 3 - ALTO

A nova classificação proposta será utilizada na elaboração do *ranking* dos países de AIBER.

13.3 Análise de Componentes Principais (PC)

Com os agrupamentos definidos e estabelecido cada país como membro de seu respectivo *cluster*, como próximo passo será executado a ANOVA para entender a variabilidade entre os grupos e auxiliar na classificação para as análises das sessões posteriores.

Principal Component Analysis: opp, waterw, immi, relig, idh, gov, epi, hpi, democ.

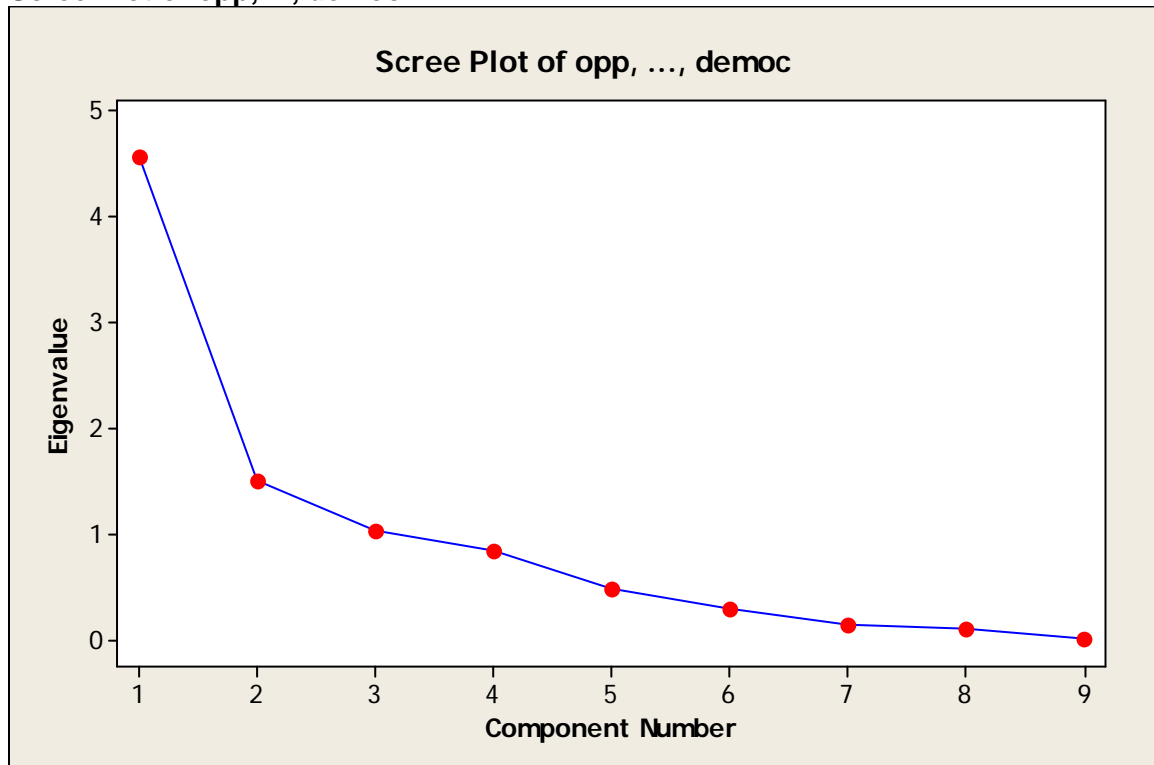
Eigenanalysis of the Correlation Matrix

Eigenvalue	4.5595	1.4955	1.0310	0.8446	0.4804	0.3004	0.1520	0.1154
Proportion	0.507	0.166	0.115	0.094	0.053	0.033	0.017	0.013
Cumulative	0.507	0.673	0.787	0.881	0.935	0.968	0.985	0.998

Eigenvalue	0.0212
Proportion	0.002
Cumulative	1.000

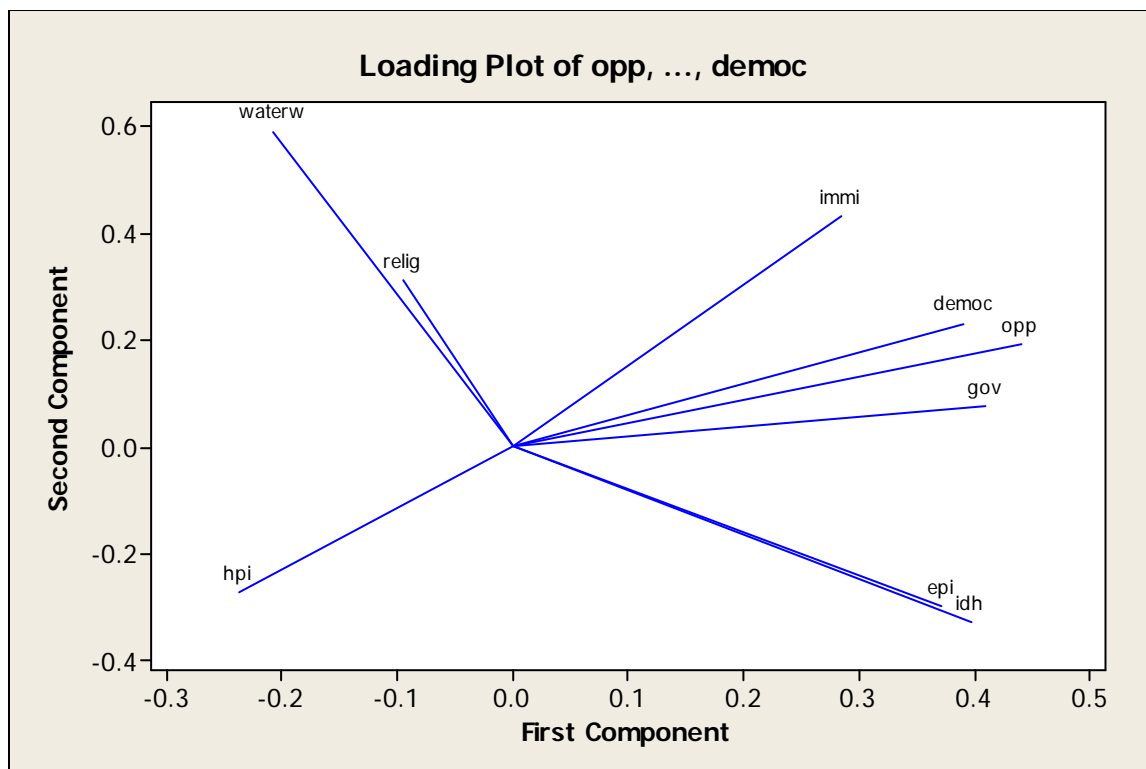
Variable	PC1	PC2	PC3	PC4
opp	0.440	0.192	0.075	-0.193
waterw	-0.207	0.591	0.377	-0.075
immi	0.284	0.431	0.155	0.380
relig	-0.095	0.311	-0.814	-0.293
idh	0.397	-0.328	-0.002	0.088
gov	0.410	0.076	-0.105	-0.389
epi	0.371	-0.300	-0.042	0.105
hpi	-0.238	-0.273	0.351	-0.658
democ	0.391	0.231	0.170	-0.351

Scree Plot of opp, ..., democ



Na análise, vemos que as componentes PC1, PC2, PC3 e PC4, explicam 88.1% da variabilidade dos dados. O que é percebido visualmente pelo gráfico acima.

Loading Plot of opp, ..., democ



As proximidades mais relevantes, encontram-se no quadrante superior direito e ficam por conta das variáveis democ (*Democracy*), opp (*Opportunity*) e gov (*Governance*), que foi demonstrado na coluna PC1 da tabela de análise de componentes. A variável immi (*Tolerance of immigrants*), se encontra no mesmo quadrante, mas um pouco mais afastada. As demais apresentam contribuição negativa ou quase nula no modelo em pelo menos um dos dois componentes (eixos).

13.4 Novo índice de performance

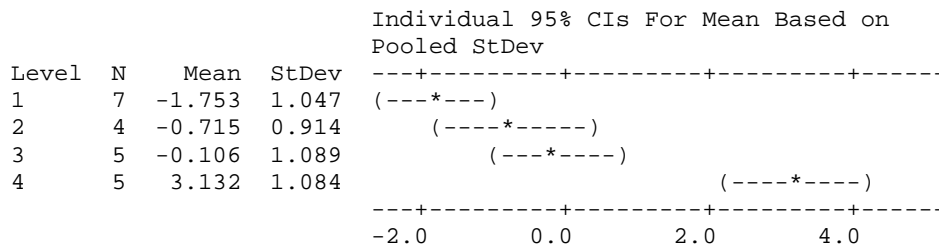
A nova PC1 criada na sessão anterior será utilizada para a criação de um novo índice de performance.

Para tanto, iniciamos executando a análise de variância ANOVA tomando como base a PC1 e a coluna de Cluster Re-Ordenados.

One-way ANOVA: pc1 versus ClusterM_reOrd

Source	DF	SS	MS	F	P
ClusterM_reOrd	3	72.66	24.22	22.22	0.000
Error	17	18.53	1.09		
Total	20	91.19			

S = 1.044 R-Sq = 79.68% R-Sq(adj) = 76.09%



Pooled StDev = 1.044

Os resultados acima, mostram pelo histograma, que manteve a mesma disposição da análise anterior (idh vs Cluster Re-Ord), porém houve aumento do coeficiente de determinação (r-quadrado), passou a 79.68% contra 58.64% descrito na sessão anterior ao ser comparado idh versus ClusterM_reOrd.

Para determinar o peso da contribuição das variáveis na PC1, será executada regressão *stepwise* utilizando como variável de interesse (*response*) a PC1 e como variáveis independentes (*predictors*), as 9 (nove) variáveis deste estudo, são elas: “opp”, “epi”, “waterw”, “hpi”, “relig”, “immi”, “gov”, “idh” e “democ”.

Stepwise Regression: pc1 versus opp, waterw, epi, gov, immi, hpi, relig, democ,...

Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15

Response is pc1 on 9 predictors, with N = 21

Step	1	2	3	4	5	6
Constant	-7.636	-10.450	-8.771	-7.469	-6.761	-7.036
opp	0.1381	0.1065	0.1098	0.1034	0.1030	0.0979
T-Value	11.93	13.01	16.75	18.69	20.32	19.00
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
epi		0.0642	0.0485	0.0484	0.0461	0.0445
T-Value		6.44	5.31	6.68	6.84	7.30
P-Value		0.000	0.000	0.000	0.000	0.000
waterw			-0.0119	-0.0113	-0.0113	-0.0132
T-Value			-3.42	-4.10	-4.50	-5.44
P-Value			0.003	0.001	0.000	0.000
hpi				-0.0136	-0.0148	-0.0121
T-Value				-3.32	-3.90	-3.35
P-Value				0.004	0.001	0.005
relig					-0.0051	-0.0048
T-Value					-2.03	-2.14

P-Value				0.061	0.050	
immi					0.0093	
T-Value					2.14	
P-Value					0.050	
S	0.752	0.425	0.336	0.267	0.244	0.219
R-Sq	88.23	96.44	97.89	98.75	99.02	99.26
R-Sq(adj)	87.61	96.04	97.52	98.44	98.69	98.94
Step	7	8	9			
Constant	-6.623	-7.036	-7.803			
opp	0.07718	0.06659	0.03030			
T-Value	8.40	7.37	*			
P-Value	0.000	0.000	*			
epi	0.04316	0.03507	0.03118			
T-Value	8.32	6.27	*			
P-Value	0.000	0.000	*			
waterw	-0.01215	-0.00927	-0.00807			
T-Value	-5.79	-4.29	*			
P-Value	0.000	0.001	*			
hpi	-0.01254	-0.01302	-0.01464			
T-Value	-4.09	-4.94	*			
P-Value	0.001	0.000	*			
relig	-0.00650	-0.00573	-0.00420			
T-Value	-3.22	-3.26	*			
P-Value	0.007	0.007	*			
immi	0.01352	0.01434	0.01750			
T-Value	3.35	4.12	*			
P-Value	0.005	0.001	*			
gov	0.01573	0.01961	0.02180			
T-Value	2.56	3.55	*			
P-Value	0.024	0.004	*			
idh		0.01748	0.03190			
T-Value		2.39	*			
P-Value		0.034	*			
democ			0.02878			
T-Value			*			
P-Value			*			
S	0.186	0.159	0.000000			
R-Sq	99.51	99.67	100.00			
R-Sq(adj)	99.24	99.44	100.00			

Com os resultados desta regressão, nota-se, analisando o coeficiente de determinação (r-quadrado), que a variável “opp” (*opportunity*) explica 88.23% da PC1 e somada com “epi” (*environmental performance index*), juntas explicam 96.44% do comportamento da variável PC1.

Ao executar a análise de regressão do PC1 e as duas variáveis de maior peso, “opp” e “epi”, temos:

Regression Analysis: pc1 versus opp, epi

The regression equation is
 $pc1 = -10.5 + 0.106 opp + 0.0642 epi$

Predictor	Coef	SE Coef	T	P
Constant	-10.4504	0.5748	-18.18	0.000
opp	0.106486	0.008183	13.01	0.000
epi	0.064230	0.009974	6.44	0.000

S = 0.424866 R-Sq = 96.4% R-Sq(adj) = 96.0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	87.941	43.970	243.59	0.000
Residual Error	18	3.249	0.181		
Total	20	91.190			

Source	DF	Seq SS
opp	1	80.455
epi	1	7.486

Unusual Observations

Obs	opp	pc1	Fit	SE Fit	Residual	St Resid
8	43.4	-0.3423	-1.2516	0.1358	0.9093	2.26R

R denotes an observation with a large standardized residual.

Que gera a seguinte equação de regressão:

$$pc1 = -10.5 + 0.106 opp + 0.0642 epi$$

Por alguma razão o modelo acima identificou a República Dominicana como uma observação “não usual”. Não houve ajustes de valor na base de dados por não ser o foco deste trabalho.

Para trabalhos futuros, este novo índice de performance será identificado como:

MeAmO = índice de performance do Meio Ambiente (epi) e de Oportunidade Social (opp).

Os países AIBER foram ordenados em ordem decrescente pelos valores do PC1 (indicador MeAmO), o qual gerou o ranking:

countries_ord	ranking	pc1_normaliz
Spain	ALTO	100.000
Portugal	ALTO	93.662
Chile	ALTO	77.400
Uruguay	ALTO	75.803
Costa Rica	ALTO	62.569
Argentina	MÉDIO	55.208
Brazil	MÉDIO	49.868
Panama	MÉDIO	39.209
Mexico	MÉDIO	36.565

Peru	MÉDIO	34.660
Dominican Republic	MÉDIO	33.901
Colombia	MÉDIO	31.318
Paraguay	MÉDIO	30.862
Bolivia	BAIXO	20.620
El Salvador	BAIXO	19.783
Ecuador	BAIXO	18.680
Venezuela	BAIXO	18.599
Cuba	BAIXO	9.896
Nicaragua	BAIXO	2.564
Honduras	BAIXO	0.045
Guatemala	BAIXO	0.000

Devido a proximidade dos resultados na análise de variância (ANOVA), agrupou-se os países dos clusters 2 e 3 que na tabela acima figuram como “MÉDIO” na escala de ranking dos países de AIBER.

Importante pontuar que a Venezuela figurava na categoria “MÉDIO” quando na análise baseada pelo IDH e passou a figurar como “BAIXO” ao efetuar as análises pelo PC1. Enquanto o México que estava como “BAIXO”, passou a “MÉDIO”, e Cuba que estava categorizada como “MÉDIO” passou a “BAIXO”. O que denota que o PC1 (Componentes Principais) conseguiu capturar um cenário um pouco distinto ao do IDH.

13.5 MAPA AIBER com ranking

Mapa com os países da região Ibero-Americana - AIBER



Legenda:

Verde escuro = ALTO

Verde médio = MÉDIO

Verde claro = BAIXO

Branco = NÃO pertencem à AIBER

Obs: A cor verde foi escolhida devido padrão do GPS que trata com a cor “verde” a sessão referente ao tema Biodiversidade – Capital Natural e Social.

13.6. CONSIDERAÇÕES

Como resultado das análises foi criado o indicador **MeAmO**, que reflete um índice de performance do Meio Ambiente (*EPI*) e de Oportunidade Social (*Opportunity*).

Cabe ressaltar que a variável *Opportunity* explica 88.23% do indicador MeAmO, e quando somada com “*EPI*” (*environmental performance index*), juntas explicam 96.44% do comportamento do indicador.

Da listagem dos países ordenados pelo indicador MeAmO, foi gerado um ranking de performance e um mapa contendo os países da região AIBER, o qual foi colorido refletindo com seus tons, o posicionamento em relação aos resultados de cada país.

Como já era esperado, países como Espanha e Portugal ficaram no topo da lista com os melhores resultados, e seguidos por estes, estão Chile, Uruguai e Costa Rica. Houveram 7 (sete) países na faixa intermediária da tabela, com destaque para Argentina, Brasil, Panamá e México. E na parte inferior da lista, com as performances mais baixas, foram destaque, Guatemala, Honduras, Nicarágua e Cuba.

14. CONSIDERAÇÕES FINAIS

Em relação ao tema Biodiversidade – Capital Natural e Social, objetivo central desta análise, percebemos no desenrolar das análises que as variáveis de maior significância apresentam um viés mais voltado ao capital social que o natural. Talvez a mensagem que está tentando emergir na evolução deste estudo é a de que o indivíduo, fazendo uma analogia à pirâmide de Maslow, necessita primeiramente tratar dos temas que impacta diretamente suas necessidades mais individuais (comunidade local, respeito e oportunidades), para só a partir daí, poder se dedicar ao entorno (proteção e preservação dos ativos naturais, qualidade de vida e humanidade).

No transcorrer das análises, uma das variáveis que mais se destacou foi *opportunity* composta pelos indicadores de Direitos individuais, Liberdades individuais, Tolerância e inclusão e Acesso à educação superior, que essencialmente demonstra o grau de liberdade dos indivíduos para exercerem seus direitos. Seguida pela variável *epi*, composta por itens como Saúde Ambiental (humana e dos ecossistemas), Impactos na Saúde, Qualidade do Ar e Nível de Água tratada e Sanitização adequada, que juntos formam o Índice de Performance Ambiental.

Com relação às análises dos países, ficou evidente o superior desempenho dos países classificados como AVECO (países de economias avançadas). Já a região Ibero-Americana (AIBER), em média apresentou resultados intermediários na maioria das análises, porém em algumas se destacou positivamente, como no caso do indicador HPI (Índice de Felicidade do Planeta).

Ainda tratando de AIBER, muitos dos resultados foram alavancados pelas presenças de Espanha e Portugal neste grupo. Retirando estes dois da amostra, houve destaque para Chile, Uruguai e Costa Rica com os resultados mais elevados. Seguidos por Argentina, Brasil, Panamá e México numa posição intermediária do *ranking*. Os piores resultados ficaram para Guatemala, Honduras e Nicarágua.

Segundo relatórios da ONU e do Banco Mundial, os países da América Latina vêm melhorando seus indicadores gradativamente, ano após ano, e mostram grande potencial para o futuro. E que parte dos resultados negativos de maneira geral, se devem ao fato de serem países “jovens”, com cerca de 500 anos de idade e baixa qualidade na educação.